ERIC Document Resume | **See Instruction Sheet** | DATE OF RESUME | DATE RECEIVED

A. PUBLICATION DATA

1. ERIC NUMBER ED002342

2. ORIGINATOR CONTROL NO(S). N.A.

3. IDENTIFICATION NO. CR 2141

DO NOT WRITE IN SPACE BELOW

MICROFILM COPY AVAILABLE YES ☒ NO ☐

COPYRIGHT GRANTED YES ☐ NO ☒

COPYRIGHT RELEASE FILED YES ☐ NO ☐

TYPE OF REPORT
PROPOSAL ☐
IN-PROGRESS ☐
FINAL ☒
FOLLOW-UP ☐

OTHER ERIC REFERENCES

ERIC SATELLITE CODE

4. AUTHOR(S): French, Joseph L. and Greer, Donald R.

5. TITLE OF DOCUMENT ABSTRACTED: Effect of Test Items Arrangement on the Physiological and Psychological Behavior in Primary School Children

6. PUBLICATION TITLE: N.A.

7. EDITOR(S): N.A.

8. PLACE OF ORIGIN
ORG. University of Missouri
ST. Columbia,
CITY, STATE Missouri
TELE. EXT.
INST/LOC CODE

9. PUBLISHER OR SOURCE Cooperative Research Program, Office of Education, Dept. of Health, Education, and Welfare, Wash., D.C.

10. VOL. N.A.

11. PAGES N.A.

12. MO./YR. 1964

13. TOTAL PP. 111

B. INDEXING DATA

14. SUBJECT AREA

15. TYPE OF ENDEAVOR Report

16. INDEXING TERMS (6-12)

		behavior, psychological		
testing	work study	behavior, physiological		
testing methods	teacher education	elementary education		
proctoring methods	student attitudes			

17. ABSTRACT (MAX. 250 WORDS)

Anxiety in the testing situation arises out of the evaluative nature of the testing procedure. Anxiety is a factor in performance. The study involved 152 first grades with IQ's between 100 and 125.

All subjects were given the Test Anxiety Scale for Children and the General Anxiety Scale for Children, after which they were divided into four groups. The Pictorial Test of Intelligence (P.T.I.) was administered to each group, including a picture vocabulary, form discrimination, information and comprehension, similarities, sizes and numbers, and immediate recall.

The apparatus used to record physiological behavior in terms of skin resistance was a Lafayette Multipurpose Polygraphy Recorder. All tests were given to all subjects by the same examiner and were given in testing rooms in the schools where the subjects were enrolled.

Findings show that test items arrangement will not be a critical factor in the behavior of similar groups of first grade children responding to similar tasks, but may be a factor for some individuals within the groups.

ERIC No.

18. SOURCE OF ABSTRACT

C. FUNDING DATA

19. SUPPORTING ORGANIZATION

20. DIVISION OR DEPT.

21. CONTRACT NO.

PROJECT NO.

22. TOTAL COST OF PROJECT

23. DURATION OR PROJECT

U.S. Department of Health, Education, and Welfare Office of Education

ERIC Form 1

ERIC

ED002342

EFFECT OF TEST ITEM ARRANGEMENT ON THE PHYSIOLOGICAL AND PSYCHOLOGICAL BEHAVIOR IN PRIMARY SCHOOL CHILDREN

Joseph L. French and **Donald R. Greer**
Principal Investigator — Research Associate

University of Missouri

MDCCCXXXIX

Cooperative Research Project 2141

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education

This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

THE EFFECT OF TEST ITEM ARRANGEMENT ON THE PHYSIOLOGICAL AND PSYCHOLOGICAL BEHAVIOR IN PRIMARY SCHOOL CHILDREN

Cooperative Research Project
No. 2141

Joseph L. French and Donald R. Greer
University of Missouri
1964

The research reported herein was supported by the Cooperative Research Program of the Office of Education, U. S. Department of Health, Education and Welfare.

ED002342

Acknowledgements

A project of this nature could not be completed without the assistance of many people. We are indebted to the people whose names appear below and hereby express out appreciation to them.

Since the subjects were the most important people in this research, we first acknowledge their cooperative spirit and especially those who wanted "to be next."

Our special thanks to the professional staff members of the Columbia Public School System whose interest and cooperation was invaluable. To the Superintendent of Schools, Dr. Robert C. Shaw, and the Assistant Superintendent of Schools, Mr. Harold E. Steere; to the principals; Mr. Burl Lowery, Dr. Paul G. Fleeman, Mr. Walter E. Bradford, Mr. Delbert Dunkin and Mr. R. Lynn Marine; and especially to the teachers, Mrs. Julia Wheeler, Mrs. Othelia Mengel, Mrs. Elizabeth Lowrance, Mrs. Sibyl Marie Kerr, Mrs. Linda Duffy, Mrs. Lois Jobe, Mrs. Eleanor Kuester, Mrs. Sara Fleming, Mrs. Clara Erickson and Mrs. Mary Jane Covington; we are deeply indebted.

To Mrs. Doni Cole, Mrs. Dixie L. Greer and to Mr. Virgil McCall, who spent long hours as operating technicians, we extend our appreciation.

We also wish to thank Dr. Robert S. Daniel, Professor of Psychology at the University of Missouri, who consulted with us concerning the theoretical aspects of physiological reaction measurement; Mr. Thomas E. Johns, Chief Engineer at the University of Missouri Instructional Television, who provided us with the necessary technical aid in equipment operation; Dr. Roy F. Kelly and Mr. Robert Patton, of the Computer Research Center, who assisted with the data processing; and Mrs. Carolyn Stange, who typed (and retyped) the manuscript. We appreciate and acknowledge their advice, counsel and effort.

J. L. F.
D. R. G.

Columbia, Missouri
August, 1964

TABLE OF CONTENTS

CHAPTER PAGE

1. Introduction 1
    Definitions 3
    Assumptions and Limitations 9
2. Related Literature 12
    Measurement of Anxiety 17
    Anxiety and Performance 24
3. Design of the Study 39
    Subjects 39
    Instruments and Apparatus 42
    Method 51
    Analysis 54
4. Results and Discussion 60
5. Summary, Conclusions and Recommendations . . . 73
    Summary 73
    Conclusions 78
    Recommendations 79

References 81

APPENDIX A. Answer Sheets for the Pictorial Test of Intelligence 87

APPENDIX B. Item Analysis Data for PTI Answer Sheet 96

APPENDIX C. PTI Behavioral Rating Scale 109

LIST OF TABLES AND ILLUSTRATIONS

TABLE		PAGE
1	Mean IQs and TASC scores for Experimental Groups	40
2	Mean IQs and TASC scores for Experimental Groups After Rematching	42
3	Summary of Analysis of Variance for PTI Means	61
4	Means, Mean Differences and t Values for Time of Administration	62
5	Means, Mean Differences and t Values for Order of Item Presentation	63
6	Summary of Analysis of Variance for Mean PTI Rating Scale Scores	64
7	Mean PTI Behavioral Rating Scale Scores for Order of Item Arrangement and Time of Testing	64
8	Summary of Analysis of Variance for GSR Means	65
9	Mean GSR Change Scores for Order of Item Arrangement and Time of Testing	65
10	Correlation Matrix between IQs and Rating Scale Scores	66
FIGURE		
1	Application of Electrodes	50

Chapter 1

Introduction

This study was undertaken as a result of a growing concern about the effects of stress generated during the intellectual assessment of children. The purpose of this study is to investigate the effects of test item arrangement on the physiological and psychological behavior of children enrolled in the first full year of school. Current techniques of intellectual assessment are many and varied but in recent years, particularly since the advent of the National Defense Education Act of 1958, standardized tests have been used more widely each year. From time to time various instruments have generated cycles of enchantment and disillusionment as their merits and faults have been discovered. This cyclical swing may be due in part to the limited knowledge of test users concerning the factors which effect the validity of scores.

All too often the criticism is raised that a given index does not properly reflect the ability of each examinee. Educators have long been aware that discrepancies exist between optimum and measured performance and some have become disenchanted with the use of tests due to discrepancies found in a few instances. Frequently tests have been used with little or no concern as to the possible factors which may influence an individual's performance in a specific testing situation.

Test developers have been cognizant of the fact that certain extraneous factors hinder the achievement of the examinee and have attempted to institute procedures which will reduce the dehabilitating effects of these factors. Of great concern to test authors are the factors which may be inherent within the testing situation itself. Some examinees may experience trepidation in a specific setting. To overcome the dehabilitating effects of this diffidence, suggestions have been offered in the literature as to how the examiner may alter his behavior in order to offset this effect. This alteration of behavior is often subsumed under "establishing rapport." The success of this procedure, however, pivots around one crucial element--the skill of the examiner in recognizing a loss or lack of rapport and his ability to do something to establish sufficient rapport to proceed.

It is not reasonable to expect that all those who use standardized tests are equally skilled or equally competent. Varying degrees of skill may be due to varying quantities of knowledge or concern on the part of some test users. They often fail to notice or have any concern about the influence of apprehension on the examinee's performance.

If, then, it is not possible to insure that rapport will be established in all testing situations, is it possible to modify the test situation so that rapport is easier to maintain? Specifically, the study reported herein is directed to determine whether a test may be altered to reduce the effects of apprehension and still properly assess the ability for which the instrument was devised.

Before proceeding further it is necessary to present certain terms which have specific meaning in this study and recur frequently enough to warrant their defining. The definitions below are based upon those presented in English and English (1958).

Definitions

Point scale - "A group of test items or problems to each of which a numerical credit is given for success." (p. 396) The score is based on the total points amassed.

Spiral omnibus test - "A test in which the various kinds of tasks are distributed throughout the test instead of being grouped...by kind and...are arranged in a recurrent pattern...of increasing difficulty." (p. 548)

Anxiety - A feeling of threat from anticipation of future harm.

Stress - Psychological force or forces which places the individual under emotional "strain or distortion in the system." (p. 529)

Tests - "A set of standard or controlled occasions for response presented to an individual with design to elicit a representative sample of behavior when meeting a given kind of environmental demand." (p. 547)

Keeping these definitions in mind, the problem of modifying a test to reduce the anxiety of the examinee may be more fully explored.

One of the most obvious ways of reducing the difficulty of a test and hence the apprehension of the examinee would be to give a test on which the individual would not encounter any difficulty. All items would be well within his ability to respond correctly. This procedure would eliminate any anxiousness which might be generated by the test but would not afford an adequate measure of his ability. It would be known that his performance was at least that of the limit of the instrument. However, such an instrument would not properly test the limits of his ability, particularly the upper limit. Such an instrument would have an inadequate "ceiling." It would not be possible to predict, with any degree of certainty, how much more could be expected and, as a result, the above procedure would not be feasible. With this possibility eliminated, the arrangement of the items of varying degrees of difficulty within the test was considered.

Item arrangement within a test follows, in the majority of instruments, one of two major patterns: increasing order of difficulty within subtest or a modification of a spiral omnibus arrangement. There seems to be little agreement among test constructors as to which method is best and the use of one or the other seems to depend primarily on the personal preference of the particular author. Either of these two methods seem to serve the purpose of assessment equally well but there is little research to support the use of either arrangement or, for that matter, any other arrangement. This finding led directly to the consideration of item arrangement in connection with the anxiousness experienced by the examinee in the test situation.

In observing test situations it has been noted that many examinees show more outward signs of anxiousness as they encounter items which are more difficult for them. As subjects approach the upper limits of their ability to correctly perform, failure occurs with increasing frequency until a point is reached when no item is successfully completed. Thus the "ceiling" is reached for that particular individual. Along with this increasing failure it has been noted that the outward signs of the examinee's discomfort also increased. Perhaps if the stress were not so great the examinee may have answered more items correctly.

If the examinee's recognized failure could be mixed with periods of recognized success, could the accompanying total amount of anxiousness be reduced also? If the items could be arranged in such a manner so the testee would experience some success throughout the entire test, rather than a period of success followed by a period of failure, perhaps anxiousness about his performance could be reduced to a level where it would not reduce his effort or accuracy of performance. If this were possible, then anxiousness and the accompanying loss of rapport could be minimized and perhaps the index of his ability derived from the test would more accurately reflect his potential performance. By arranging the items so that probable success would surround those items on which probable failure would occur, it would be possible to assure that some success would be encountered within any given group of items throughout the entire length of the test. As this order of arrangement was considered in detail other possibilities were suggested.

Clinical experience in testing situations again served as a guide-line. Some individuals seemed to be more anxious with regard to certain types of tasks than others. In the traditional subtest arrangement, items of the same kind are placed together and in increasing order of difficulty.

While this has the advantage of allowing the testee to go from easy items to difficult items and then back to easy items throughout the test, some examinees seemed to express more anxiousness during a numerical subtest or those requiring skills of perceptual organization than during other subtests, while others seemed to express more anxiousness during highly verbal subtests. The spiral omnibus order of item arrangement reduced the length of time an examinee would be exposed to a type of task in which he has some expectations of failure.

Two points of consideration for the arrangement of the items within a test have been described; one dealing with the placement of items to assure some feeling of success at given times throughout the test and the other dealing with the heterogenous arrangement of items so as to reduce the concurrent time that the individual will be exposed to any particular type of task within the testing situation. The present study is concerned with these two points. The investigation deals specifically with the effects on subject behavior of four different arrangements of items within a new point scale designed to measure intelligence. The four arrangements are:

1. Published order - items arranged approximately in increasing order of difficulty within subtests.

2. Absolute order - items arranged in increasing order of difficulty within subtests according to the standardization data.

3. Random order - items placed randomly throughout the test without regard to subtest or item difficulty.

4. Patterned order - items placed in a sequence of two easy, one hard throughout the test without regard to subtest.

With these four arrangements the effects of subtest and item arrangement may be studied. The general hypothesis is that no difference will be found in intelligence scores as measured by a point scale, in behavior as observed and recorded by an examiner when the items are presented in the different orders, or in physiological reaction as measured by a galvanometer.

The assumptions and limitations which have been accepted in connection with this study should be stated at this point.

Assumptions and Limitations

One of the first assumptions that was made in this study dealt with the function of anxiety in relation to the testing situation. It was assumed that there was a certain amount of test anxiety present in all individuals in the testing situation. It was further assumed that such anxiety would be related to test performance. It was also assumed that if anxiety does exist in some degree it could be quantified and some index be derived to indicate its presence or absence in the individual at any given time.

Other assumptions concerning the test instruments were made. A major assumption was that each item contributed the same weight to the total score regardless of where it was placed in the order of items. An item was assumed to measure that portion of behavior completely unrelated to the sequence where it was placed. This assumption is common to all point-scale tests. The item will reflect the same amount of contribution to the total assessment as any other item regardless of the difficulty of the item. An easy item for the individual contributes as much to his total score as does a difficult item. The value of an item as a unit of measure is completely unrelated to the other items within the test except as they are summed to give a total score.

It was also assumed that the difficulty of the items was unrelated to the sequence of items in the test and that item difficulty as determined under standardization procedures was an adequate reflection of the difficulty of that item for any subsample of the population in general. This assumption basically is one of sampling and involves two parts. One, that the standardization sample was truely representative of the population as a whole; and two, that any subsample drawn from the total population will not differ in any significant manner from the population as a whole and, therefore, not from the standardization population. In this circuitous manner we are assuming that an item would be as difficult for any sample drawn from the general population as it was for the standardization sample. Since this can only be determined subsequently it must be assumed antecedently.

The above assumptions were necessary before any study could proceed and, therefore, were not considered to be a deterent to the investigation. However, it is fully realized that certain limitations must be imposed on the results. For this reason certain limitations were inherent within the study and they will be listed below.

Perhaps the most important limitation of any study is that of the representativeness of the sample. The results of this study can be applied only to groups similar to the particular group with which the study was conducted. Only in those characteristics which the group does not significantly differ from the general population can there be any generalizations made from this study.

There is a limitation dealing with the generalization from the instrument used for assessment of intellectual ability. It was necessary to choose a point scale type of test in order to satisfy some of the above assumptions. The findings in regard to this point scale can only be limited to the particular scale and it is not intended that it be representative of all point scales of intellectual assessment. The findings may generally be hypothesized for tests of the same type in as far as they satisfy the above assumptions as to independence of items. However, similar studies should be conducted verifying the hypothesis.

A futher limitation is that all examiners may not be as familiar with the instruments and apparatus which were used in this study as the investigators who conducted this study.

In this chapter the need for the study has been established and the problem has been identified. In the following chapter a theoretical position will be proposed and the literature reporting studies involving the methods of measurement of anxiety and those reporting the relationship between anxiety and task performance will be reviewed.

Chapter 2

Related Literature

In recent years a great deal of research has been conducted concerning the relationship between anxiety and various types of performances. As a result of this increased interest in the relationship of anxiety and performance, particularly intellectual functioning, some concern has been expressed about the nature and origin of anxiety. In order to bring the important factors into clearer focus, an attempt will be made to summarize the literature dealing with anxiety in testing situations.

Most investigators feel that in order to adequately define anxiety it is first necessary to separate anxiety from other emotional states which are present in the organism from time to time. One of the emotional states which is often considered closely related to anxiety is fear. Often these two concepts are used interchangably but Leavitt (1957), in his review of the topic, takes a very different position. He makes a clear differentiation between fear and anxiety. Fear is an emotional state incurred when a certainty exists that danger is a present threat to the organism. Anxiety on the other hand can occur only when an uncertainty exists as to whether "suffering" will occur. Later he states that what the conscious ego preceives as anxiety "seems to be an expression of uncertainty by the unconscious portion of the ego as to whether the repression mechanism and/or other ego defenses will hold fast to prevent an instinct presentation from reaching consciousness." (1957, p. 70)

Leavitt's position is one which would generally be accepted by personality theorists but there is some disagreement between the personality theorists and the S-R theorists as to the nature of anxiety. Whiteman (1957) believes this schism exists only because of different ways of viewing the same mechanism. He points out that the S-R theorist stresses the relationship of signals of noxious stimulation to anxiety whereas the personality theorist views anxiety as a reaction to threat, danger, or trauma. Whiteman attempts to close this schism by taking the position that anxiety is a function of frustration; frustration which occurs because of what he terms an "adaptive lack" on the part of the organism under varying states of drive. This position should satisfy both groups of theorists in that it shows the relation to drive as espoused by the S-R theorists and is closely related to the position taken in Leavitt's review. The "adaptive lack" would occur because there is not a clear certainty that "suffering" will occur; hence, only a vague threat to the organism.

Since the relation of anxiety as it is evidenced in the testing situation is the focal point of the present study, a brief consideration of current test anxiety theory will be presented here.

Sarason, S. B., Davidson, D., Lighthall, Waite, and Ruebush (1960), after a study of test anxiety in children, felt that anxiety is generated in the child because of the evaluative nature of the situation which exists between the child and his parents. During the child's growing years, he is constantly being evaluated by his parents. This evaluation gives rise to feelings of either hostility or aggression on the part of the child toward his parents. Because of his subservent position to his parents, he is not allowed to give free expression to these feelings. In order to avoid either physical or psychological punishment he tends to repress these feelings. The psychoanalitic position holds that repressed material threatens to break through to the conscious level. Thus, the uncertainty of the "suffering" stated by Leavitt exists. Stanley-Jones (1957) tends to support this concept. He states that "anxiety is quantitatively due to aggression that has been repressed...."

The theoretical position of Sarason, S. B., et al., and Stanley-Jones receives some emperical support from a study conducted by Learmonth, Ackerly, and Kaplin (1957). They conclude that: "...response to stress is correlated positively with a group of personality variables that have in common restraint and curtailment of unpleasant or prohibited feelings or actions." (p. 157)

According to the theoretical position above, the testing situation has an evaluative nature and, therefore, feelings of hostility or aggression arise. These feelings are repressed and, hence, anxiety is generated.

By extrapolating from this orientation, some interesting possibilities arise. If anxiety arises because of an "adaptive lack" or because an uncertainty exists as to the possible "suffering" on the part of the organism, it would seem that the more complex the task on which the evaluation is made the greater the anxiety. The more complex the task the greater the possibility that there will not be a pattern which will lead to the successful performance of the task within the behavior repertory of the organism. There will exist the possibility of Whiteman's "adaptive lack." Because it is only a possibility, the element of uncertainty may occur and thus Leavitt's uncertainty of suffering may occur.

One other interesting speculation may be made. This deals with the relationship between ego involvement and anxiety. The greater the ego involvement on the part of the individual the greater the possibility of "suffering." If his behavior is nonadaptive, the more his ego integrity may be threatened. With such threat, the greater may be his resentment, the greater may be the repression and thus the greater may be the anxiety.

With the theoretical position established that the testing situation may generate anxiety within the individual, the areas of anxiety measurement and anxiety and performance will be reviewed.

The subjects of "anxiety" and "measurement" are found in great abundance in the literature. While most of these articles could be related to the current study in some manner, an inclusion of a majority of the research reports would serve only to obscure the purpose of this review. The main concern in the current investigation is the measurement of anxiety and the relationship between anxiety and the testing situation. Recently Merwin and Gardner (1962) reviewed most of the recent literature concerned with test performance and anxiety and Martin (1961) presented an excellent review of the research dealing with anxiety measurement. Although these two reviews present excellent summaries of the research in the two areas, the needs of the current study would be better served if a closer look were taken at some of the more pertinent studies. The specific studies will be grouped into various catagories to facilitate analysis. Following the lead of Merwin and Gardner (1962) and Martin (1961), the two major divisions will be the measurement of anxiety and the relationship of anxiety and performance. Within these two major divisions will be sub-divisions. The first will contain physiological measures of anxiety and psychological measures of anxiety. The second major division will contain sub-divisions pertaining to anxiety and verbal performance, anxiety and achievement and others.

Measurement of Anxiety

A great variety of physiological varibles have been measured in connection with anxiety and anxiety states. Martin (1961) reported in his review that systolic blood pressure, diastolic blood pressure, heart rate, cardiac output, hand temperature, palmar conductance, respiration rate, frontalis muscle tension, finger tremor and salivary output have all been used by investigators inquiring into the measurement of anxiety. Martin, in the same review, reporting on a theory proposed by Funkenstein, et al., states that they suggest that the physiological reaction accompanying anger directed outwardly is a nonepinephrine-like reaction while the reaction accompanying anxiety is a epinephrine-like reaction. Reviewing the studies conducted to investigate the physiological reactions accompanying injections of epinephrine and nonepinephrine, Martin found that, in general, epinephrine leads to increased palmar conductance, systolic blood pressure, heart rate, cardiac output, forehead temperature, central system stimulation and blood sugar level. Decreases were noted in diastolic blood pressure, peripheral resistance, hand temperature, and salivary output.

Although a great number of physiological measures may adequately reflect anxiety, practical and ethical considerations of the current study made it necessary to restrict the selection of a measure to those which would be relatively easy to secure, would cause the least amount of discomfort to the subjects involved in the study, and would be relatively uneffected by the erratic movements which are characteristic of the subjects in the study. For these reasons, a palmar sweating measure was decided upon as a physiological index of anxiety. Since this will be the measure used, the literature reviewed will be primarily restricted to studies involving some index of palmar sweating.

Palmar sweating is measured with a variety of techniques and given many lables. In order to introduce some clarity, all references to palmar sweating will be referred to as the galvanic skin response (GSR) and the specific measures involved in each study will be noted.

In most of the studies in which the GSR has been used, some outside index of anxiety has been used concurrently.

These indicies are generally either a scale designed to measure reported anxiety or clinical judgement of anxiety. One other procedure which has been used is to assign the subjects tasks which have been judged to induce anxiety. Two studies which use the Palmar Prespiration Index (PPI) as a physiological measure of anxiety and the Taylor Manifest Anxiety Scale (MAS) as a criterion index, found no relationship between the two measures. Lotsof and Downing (1956) and Calvin, McGuigan, Tyrell, and Soyars (1956) reported no significant relationship between the two measures in the experiments which they conducted. Two other studies using the Palmar Sweat Index (PSI), conducted by Bixenstein (1955) and Rosenstein (1960) reported quite different results. The former investigator found a significant relationship between the density of the PSI and reported periods of stress. Rosenstein found essentially the same phenomena. The PSI showed significantly greater increments for subjects who were placed under stress by instructions from the investigator. In the same study the Test Anxiety Scale (TAS) was also used and no relationship was found between the two measures.

The findings reported in the studies above indicate that the psychometric measures were not significantly related to the measures PPI or PSI. Quite different results were reported when other physiological measures were compaired with psychometric measures.

McDonnell and Carpenter (1959) found that the skin conductance level was significantly related to scores received on the General Anxiety Scale (GAS). The same results were also reported by McDonnell and Carpenter (1960) in another study they conducted. Raphelson (1957) used the Test Anxiety Questionnaire (TAQ), the MAS and a need achievement scale and found that subjects who were considered relatively anxious (low need achievement scores and high TAQ scores) showed an increase in skin conductance while the group classified as nonanxious showed a decrease in conductance. These differences were significant but there were no significant results when the MAS was used to classify the subjects. Raphelson and Moulton (1958) replicated the aforementioned study with the exception that the subjects had no reference as to the accuracy of their performances. In this study there were no differences between conductance levels in a neutral task and the task designed to induce stress.

Berry and Martin (1957) investigating the relationship between TAS scores and conditioning of the sharp increases in skin conductance found that there was no relationship between conditioning and reported anxiety level. In this study girls were found to be more reactive than boys. Gilberstadt and Davenport (1960) carried out a conditioning experiment of the sharp increases of conductance and found that the conditioning was different for subjects divided into three groups by clinical judgement but no differences were found when the subjects were grouped according to MAS scores. Bitterman and Holtzman (1952) obtained results which support those of Gilberstadt and Davenport, thus indicating that the clinical index seems to be a better predictor of conditioning than scores obtained from the MAS.

A number of studies using a variety of GSR measures may be broadly classified under the heading of "task induced anxiety." Novak, Moriarty, Koltuv, and Welsh (1956), established avoidance and nonavoidance conditions in a reaction time study. They found that two of the three nonavoidance groups showed lower ohmic resistance than their prestimulus levels but no such differences were found for the avoidance groups.

Using a group of hospitalized male patients, Sines (1957) divided them into three groups on the basis of psychological test and/or interviews. Stimulus cards were selected which would bring into focus the area of conflict for these three groups. The results indicated that the stimuli evoked changes in physiological activity which were statistically significant. The one measure of physiological activity used in this study was GSR.

Percent of conductance change was found to be positively related to the difficulty and complexity of motor tasks in a study conducted by Venables (1956). For the normal subjects used in the study, a decrease in the difficulty of the task was followed by a decrease in conductance but this was not the case for the subjects classified as psychoneurotic.

A continuous mental task was used by Du Tiot (1956). He reported an over-all increase in skin conductance in the majority of the records obtained.

Martin (1959) obtained continuous skin conductance levels during a stressful interview. The subjects took a feeling inventory with three different response sets. There was a significant correlation only for the groups which were instructed to respond to the inventory in terms of how they felt in the last fifteen minutes. There was no significant relationship between the skin conductance and MAS scores.

An Autonomic Perception Questionnaire was used by Mandler, Mandler, and Uviller (1958). The subjects were given three stressor tasks after having been divided into two groups on the basis of the scores received on the questionnaire. Heart rate, pyschogalvanic responses (PGR), respiration, face temperature, and blood volume measures were taken throughout the tasks. The differences between the two groups were significant for heart rate, PGR, respiration and total reactivity scores.

Five learning sessions for three groups of fourth graders divided on the basis of intelligence scores were used by Ringness (1962). The tasks assigned to each group were kept just above the current ability level for all subjects. Five indicies of GSR were obtained throughout the five learning sessions. These were: mean base level, slope of line, varibility about the regression line, combined height of all peaks of 500 ohms or more, and the number of such peaks. The base levels were seen to decrease with each succeeding task. Sex differences were found for base levels and initial arousal levels. Differences between the groups were found only in regard to base level and slope of base level.

Kimmel and Hill (1961), Gordon, Martin, and Lundy (1959), also found decreases in skin conductance concomitantly with periods of stress that were induced in the subject.

The literature reviewed in connection with physiological measures of anxiety indicate that while some measures of GSR, particularly the PPI and PSI, do not adequately reflect the anxiety present; other measures, particularly skin conductance and skin resistance, seem to consistantly reflect this state. In the next section of the review of the measurement of anxiety, the psychological measures will be given closer attention.

A number of the instruments which have been designed to be psychometric measures of anxiety have been mentioned in the section on physiological measures of anxiety. Those mentioned were: the Taylor Manifest Anxiety Scale (MAS), the Test Anxiety Questionnaire (TAQ), and the General Anxiety Scale (GAS). Others which will receive mention in this review are: the Childrens' Manifest Anxiety Scale (CMAS), Bendig's revision of the MAS, Heineman's revised MAS, the Childrens' Anxiety Pictures, and the IPAT Anxiety Scale.

The scales mentioned above do not constitute a complete list of the instruments which have been used in the investigations reported in this review. Since a description of these instruments is available elsewhere no attempt will be made to describe them here. Rather than engage in a description of scales, a review of the studies in which these instruments have been used would seem to be more fruitful. Therefore, a review of the literature of the effects of anxiety on performance will follow.

Anxiety and Performance

The relationship of anxiety to verbal performance has been studied by a number of researchers. Barnard, Zimbardo, and Sarason, S. B., (1961) investigated the relationship between anxiety and verbal behavior in third grade children. Matching on intelligence test scores, sex and grade, the subjects were divided into three groups according to scores derived from an anxiety measure. The subjects were again divided into two groups and interviewed under authoritarian and permissive conditions respectively. The high anxious subjects made more references to negative affect under the authoritarian setting than they did under permissive conditions. High anxious males had less forceful voice quality and all high anxious subjects tended to give more concrete answers than low anxious subjects. The high anxious subjects' responses tended to be more incomprehensible than those in the low anxious group.

Sullivan and Calvin (1957), investigating the relation of spoken and written vocabulary to a standardized vocabulary test, intelligence and anxiety, found that the amount of uncommon words spoken during an interview was significantly correlated in the negative direction to the score obtained on the MAS.

Studying the effects of anxiety and situational stress on communicative efficiency, Gynther (1957) found that subjects having low scores on the Welch A Scale were significantly higher on the measure of communicative efficiency than the subjects classified as highly anxious.

Rawn (1958), using the IPAT Anxiety Scale and Grace's Hostility Inventory, found that subjects who were least prone to verbal indication of physical hostility were most prone to show the covert anxiety, while those most prone to verbal expression of physical hostility had the greatest overt anxiety.

The above studies indicate that anxiety may be significantly related to the quality, efficiency, and kind of verbal behavior evidenced by a particular group of subjects and that the degree of influence seems to be an inverse function.

The relationship of anxiety and achievement has been studied frequently. Alpert and Haber (1960) investigated anxiety in academic situations. Using a variety of anxiety measures, they found that specific measures of anxiety more often correlated significantly with the indices of academic performance of college freshmen than did the more general measures.

Studying anxiety, intelligence, and achievement of children divided into three intelligence groups, Feldhusen and Klausmeier (1962) found that the correlations between measured anxiety, intelligence test scores, and achievement test scores were all significant. An interesting finding was that there seemed to be no relation between anxiety and achievement test scores for the group with high intelligence test scores. The authors explained this on the basis that the more intelligent subject may be able to clearly assess the real and present danger inherent in academic testing situations and, therefore, were not threatened by it.

McCandless and Castenada (1956), studying the relationship between anxiety, school achievement, and intelligence in fourth, fifth, and sixth grade children, found that the more complicated skills such as reading and arithmetic suffered more interference from anxiety than did the simpler mnemonic skills and that the interference was greater for girls than for boys.

Sarason, I. G. (1957b), using a group of college undergraduates as subjects, found the performance on academic achievement measures of a group classified as low anxious on the Test Anxiety Scale (TAS) to be superior to that of a group of subjects classified as high anxious. Also the group lowest on the anxiety measure had significantly better grade point averages than the other group used in the study.

Another study using grade point averages as one index of academic achievement was that of Spielberger and Katzenmeyer (1959). These investigators found, after dividing their sample of college juniors into three groups on the American College Examination (ACE) and the MAS, that grade point average was unrelated to the anxiety measure for the high and low anxiety groups but significantly related to the grade point averages of the middle anxiety group. The investigators explained this finding on the basis that college work was too difficult for the low ability student and that high ability students tended to do well regardless of the anxiety level. This finding seems to support that of Feldhusen and Klausmeier which was reported above.

Pickrel (1958), using United States Air Force Basic Airmen as subjects, divided them into two groups on the basis of a modified MAS and administered four tests of different types. On the scales measuring the more complex activities, the nonanxious subjects scored significantly higher than the high anxious subjects. This finding is in keeping with that of McCandless and Castenada which was cited above. Anxiety as such seems to effect adults much in the same manner in complex tasks as it does children.

Grooms and Endler (1960) trichotomized a group of male college students on the TAQ and collected academic aptitude test scores and semester, cumulative, and predicted grade point averages. They found a significant negative correlation between TAQ scores and the scores derived from the measure of academic aptitude but no significant relationship between the scores on the TAQ and the grade point averages. At first glance this finding appears to be at variance with those of Spielberger and Katzenmeyer (1959) but they used a general anxiety scale whereas Grooms and Endler used a scale specifically designed to measure anxiety in the testing situation.

Lawrence (1962) collected scores on MAS, a projective measure of achievement motivation, and the ACE from a group of college students. One half of the group were told that their performance on the ACE was important while the other half were told that the test was given only for the purpose of preparing norms. The group designated as high anxious on the anxiety scale scored significantly lower on the ACE when the test was viewed as important. When the test was viewed as unimportant, anxiety was not related to performance. The investigators contend that stressful situations seem to interfere with successful performance of high anxious subjects on a complex task.

Atkinson and Litwin (1960), using TAS scores as an index of motive to avoid failure, collected scores on two need achievement scales, distances from a goal in a ring toss game, time taken on a course final and the score achieved on the final. They found that subjects whose need to avoid failure was greater than their motive to achieve showed less efficiency or accomplishment than those subjects whose need to avoid failure was less than their need to achieve.

The research reviewed dealing with anxiety and achievement either in academic settings or on tests of academic aptitude seems to indicate that anxiety is inversely related to academic performance indices such as grade point average and achievement test performance. Anxiety seems to be dehabilitating in complex situations but the dehabilitating effect is not as great among people of superior intelligence as among those of lesser intelligence or among those possessing a higher than average need to achieve.

The relationship between anxiety and performance in the learning situation has received a great deal of study. The learning situations have been varied and the tasks have ranged from simple to rather complex. From the studies in which the MAS was used (Buchwald, 1959; L'Abate, 1959; Sarason, I. G., 1957a; Sarason, I. G., 1956; Sinha and Singh, 1959; Stevenson and Iscoe, 1956), from the studies in which the TAS was used (Sarason, I. G., 1961; Sarason, I. G., 1958; Smock, 1956; Wiener, 1959), and from other recent studies (Silverman and Blitz, 1956; Spence, Farber and McFann, 1956; Castaneda, Palermo, and McCandless 1956; Waite, R. R., Sarason, S. B., Lighthall, Frederick, and Davidson, K., 1958) it has been concluded that anxiety seems to effect the rate of learning, the number of errors committed, and the time necessary to make discriminations. The effects of anxiety seem to be mediated by the type of instructions, complexity of the task, and the progression from easy to difficult items within the task. In some studies the effects of anxiety seem to be different for males than for females.

The investigation of performance on numerical tasks and anxiety seems to be somewhat limited but a few do exist. Callahan and Keller (1957) used the Childrens' Anxiety Pictures, Goodstein and Farber (1957) used the MAS, Hafner, Pollie, and Wapner (1960) used the CMAS, and Sarason, I. G., (1957b) used the TAQ.

In most of the studies dealing with anxiety and numerical performance, support can be found for the position that anxiety and superior performance on numerical tasks are inversely related.

In a study of reaction time, Castaneda (1956) found that response amplitude of high anxious subjects was greater than for low anxious subjects. The speed of reaction, however, was higher for the low anxious group.

Davidson, W. I., Andrews and Ross (1956) studied the effects of stress and anxiety on high-speed color naming. Having three levels of tasks stress, three levels of failure stress and two levels of anxiety, they found that anxiety served to increase the subjects' susceptibility to the effects of stressful conditions.

Shepard and Abbey (1958) investigated performance on a complex motor skill task as effected by manifest anxiety.

They found that the low anxiety subjects were superior on all tasks to the high anxiety subjects.

Anxiety seems to effect speed of reaction of perceptual motor skills and overall performance on these tasks. As the complexity of the task is increased, the effects of anxiety seem also to increase.

Some investigation has been carried out to make inquiry into the effects of anxiety on abstract concepts. Lipman and Griffith (1960) explored the relationship between anxiety and concept formation with institutional retardates as subjects. Using scores derived from the CMAS to designate the two anxiety groups, the investigators administered a concept formation task within which the subjects were required to report similarities within sets of words. As the abstractions became more difficult, the high anxious subjects' performance was significantly poorer than the low anxious subjects'.

Kerrick (1956) divided her group of high school subjects on the basis of upper and lower quartiles of an intelligence test score distribution. These subjects were then divided on the basis of upper and lower quartiles of the MAS score distribution. With four groups defined, a task consisting of judging a concept against a series of discriptive scales was undertaken by all subjects. The finding indicated an interaction exists between intelligence and anxiety.

With increased anxiety, high intelligence groups tend to become less discriminating while the low intelligence groups tend to become more cautious.

Increased anxiety seems to have an adverse effect on concept formation and concept judgement. The effects of anxiety seems to increase with an increase in the complexity of the task.

A study concerned with perception and anxiety and one dealing with the function of anxiety on a complex reasoning task will provide some interesting results. The study involving perception was carried out by Ruebush (1960). He divided sixth grade boys into three groups on intelligence test scores, into two groups on _Yale Test Anxiety Scale_ scores, and into two treatment groups. The task was to find an imbedded figure. The two treatments consisted of administering the items in an order from easy to difficult and the second order was to administer the items from difficult to easy. Three scores were derived: latency, number of correct solutions, and a cautiousness index. It was found that high anxious subjects obtained higher cautiousness scores. High cautious subjects did better than low cautious subjects and the performance of high anxious subjects was superior to low anxious subjects at the low and middle intelligence range. The high anxious subject tends to be more cautious and in situations where caution is an important factor tends to do better.

Anxiety and reasoning tasks were investigated by Buskirk (1961). Using college students as subjects, he administered three measures of manifest anxiety, two reasoning tasks, a logical deduction task, and a figure analogy task. The high manifest anxiety groups performed at a superior level on the complex logical deduction task. There were no differences between the groups on the other tasks. The investigator felt that the results of the study supported the interpretation that high anxious subjects have a stronger desire to do well as long as the situation does not present a serious threat to the ego.

The authors of the study dealing with perception and the one dealing with reasoning provide support for the contention that a high anxiety level may be facilitating in some instances provided that the situation is such that the performance called for does not involve elements which would lead to a threat to the subject.

The studies reviewed to this point have been mostly concerned with one type of performance in relation to anxiety. One situation which incorporates many of these specif-types of behavior is that of intelligence testing. The relationship of anxiety to performance on an intelligence test is the core of this present study and, therefore, a few studies will be reviewed.

There have been a number of studies conducted to examine the interrelationship of anxiety and intelligence test performance. Hafner, Pollie, and Wapner (1960), administered the CMAS and the *Weschler Intelligence Scale for Children* (WISC) to subjects ages 10 through 12. There were no significant correlations between the intelligence quotients derived from the full scale and the verbal and performance subtest. Of the subscales, only block design and coding were significantly negatively correlated with the anxiety scores. In the total study, however, the authors report a negative correlation between manifest anxiety and WISC functioning.

The *Otis Quick Scoring Mental Abilities Test* was used by Keller and Rowley (1962). The Otis Test, the CMAS and the appropriate *Metropolitan Achievement Test* were administered to fourth, fifth and sixth grade children. The analysis indicated that anxiety and intelligence scores had a significant negative correlation only for sixth grade girls. That anxiety, as reflected by the particular scale used in the study, develops only after a number of years in the school setting was suggested but not supported by Zweibelson (1960) who used the *Otis Alpha*, *Otis Beta*, and the *Davis Eells Games* in his investigation of fourth grade school children. He administered the TAQ and found that low anxious subjects do better than high anxious subjects regardless of the type of test used. Zweibelson found significantly greater negative correlations between the Otis test scores and TAS scores than between Davis-Eells scores and TAS scores. He indicated that the game-like quality of the Davis-Eells Games probably accounted for the lower negative correlation between the intelligence scores derived from that instrument and test anxiety scores.

Lighthall, Ruebush, Sarason, S. B., and Zweibelson (1959) used the Otis Beta and the Davis-Eells Games in their study of anxiety and mental ability. The Test Anxiety Scale for Children (TASC) was given to children in the fourth grade and the two intelligence measures were administered at the beginning of the fifth and again at the beginning of the seventh grade. Gain scores and mean scores for the two intelligence measures were analyzed. The subjects were divided at the median on the anxiety scale distribution. The high anxiety group was lower in terms of gain scores but the difference was not significant. When the mean scores on the two intelligence measures were compared as a function of anxiety, a significant effect of anxiety on performance was noted. The low anxiety group showed more gain on the Otis Beta while the high anxiety group showed more gain on the Davis-Eells Games.

Phillips (1962) administered the California Test of Mental Maturity along with achievement measures. Using the CMAS as a measure of anxiety, he found that high anxious subjects had lower intelligence scores than the low anxious group.

The Weschler Adult Intelligence Scale (WAIS), Raven's Progressive Matrices, the Bender-Gestalt and the Bender-Gestalt recall test were used by Siegman (1956). The MAS was administered to 35 medical and psychiatric patients along with the measures of intelligence. Correlations were obtained between the various measures. The low anxiety subjects' performance was superior to the high anxiety subjects' on Raven's Progressive Matrices, the Bender-Gestalt recall task and the timed sections of the WAIS. There were no significant differences in performance on the untimed sections of the WAIS or the Bender-Gestalt.

Performance on intelligence tests seems to be inversely related to anxiety. The influence of anxiety may be mediated by the format of the test and may be more of a factor when time limits are established.

Although there may be many other areas of performance which may be influenced by anxiety, the review of this section touches only those areas which were closely related to types of subject behavior required in the testing situation, particularly intelligence testing. In the studies reviewed, it was indicated that anxiety may be a significant factor in verbal performance, achievement, learning, performance on numerical tasks, perceptual motor skills, concept formation, reasoning, and intelligence testing.

This chapter has presented a theoretical position on the origin of anxiety in the testing situation, a review of literature on the measurement of anxiety and a review of the literature on the effects of anxiety on performance. In the next chapter, the design of the study will be presented.

Chapter 3

Design of the Study

Subjects

The subjects for this study consisted of 160 pupils who were enrolled in the first grade of five public schools in Columbia, Missouri. To enter the first grade in Columbia, children must be six years of age by September 15th, therefore, during the eight months of data collection the subjects were six and seven years of age.

From the children enrolled in the schools, those with *California Test of Mental Maturity* (CTMM) total IQs between 100 and 125 were selected for this study. Most children responded to the CTMM during May of their kindergarten year. The exceptions were tested during the following September.

All subjects selected for this project were given an exposure to the testing situation in which a GSR recorder was attached and in operation prior to the administration of the intelligence test used in the study. While the recorder was attached, the *Test Anxiety Scale for Children* (TASC) and the *General Anxiety Scale for Children* (GASC) (Sarason, S. B., et al. 1960) were administered verbally by the examiner. This procedure served two purposes. First, it exposed the subjects to the elements of the testing sessions which were to follow. Each child was allowed to respond to the examiner and to become familiar with the equipment. The procedure was employed to minimize the anxiety which might have been artificially imposed because of the novelity of the situation. Second, the TASC scores from the pretrial exposure provided a criterion for matching the experimental groups.

After all subjects were administered the TASC and the GASC, they were divided into four treatment groups of 40 subjects each and designated as Group I, II, III, and IV. Each group was composed of 20 boys and 20 girls whose mean CTMM IQs and mean TASC scores were very similar. The means may be found in Table 1. The appropriate t test was performed to determine if the differences observed between means were within chance fluctuations and the hypothesis of no significant differences was supported.

Table 1

Mean IQs and TASC scores for Experimental Group

(N = 40 in each group)

	CTMM	TASC
Group I	$\bar{X}$ 113.150	$\bar{X}$ 9.775
Group II	$\bar{X}$ 113.150	$\bar{X}$ 9.750
Group III	$\bar{X}$ 113.200	$\bar{X}$ 9.800
Group IV	$\bar{X}$ 113.150	$\bar{X}$ 9.800

The groups were thus matched on three variables: sex, IQs, and anxiety scores. These three variables were selected in order to control those factors which were observed to be related to anxiety in test situations described in the literature. The range of IQs from 100 to 125 was selected for two reasons. The first was to insure that all subjects would have some success on the individual intelligence test but none would have complete success, and the second was that the average or high average range of intelligence would provide a more stringent test of the hypothesis in the study since test anxiety does not seem to be as high in this range as others.

During the course of the investigation, five subjects were unable to complete the entire series of tests. Two subjects were lost to the study from both Group III and IV, and one was lost from Group II.

Fortunately a boy and a girl were lost from Group III and Group IV. This made necessary a rematching of the groups on the three original variables, so that the final data analysis was conducted on groups of 38 each, 19 boys and 19 girls. After eliminating two subjects from Group I and one from Group II, mean IQs and mean TASC scores were recomputed for each group and the appropriate t test was carried out. There were no significant differences found between the means on either variable. The mean CTMM IQs and the mean TASC scores for the subjects used in the study are found in Table 2.

Table 2

Mean IQs and TASC Scores for Experimental Groups After Rematching

(N = 38 in each group)

	CTMM	TASC
Group I	113.816	9.868
Group II	112.736	9.368
Group III	113.816	9.763
Group IV	113.0263	9.711

Instruments and Apparatus

The *Pictorial Test of Intelligence* (PTI) was the point scale used in this study. The PTI is an individual test of general ability designed for administration by an examiner to one child at a time. It is a revision of the *North Central Individual Test of Mental Abilities* and was recently published by Houghton Mifflin Company. Questions are presented orally by the examiner in conjuction with large (11"x 11") picture cards on which are represented four possible choices or answers to the examiner's request. The child responds to the verbal instructions of the examiner by selecting one of four drawings on each card.

The test has no time limits and has 160 items arranged in six subtests. The subtests are: Picture Vocabulary, Form Discrimination, Information and Comprehension, Similarities, Size and Number, and Immediate Recall. While the subtests provide for the measurement of several different facets of mental functioning all subtests appear to be heavily loaded with "g."

The subject may make a response to the verbally posed requests of the examiner by pointing to the response card drawing which he considers to be correct.

A full description of the standardization data and the test items is found in the manual which accompanies the test (French, 1964). In order to acquaint the reader with the format of the test, a few sample items from each subtest are listed below:

Picture Vocabulary

V 1 Find the picture of the chair.

V 29 Find acquiring

Form Discrimination

F 2 Find the one that looks just like this.

F 24 Now find the one that can be finished to look just like this one.

Information and Comprehension

I 1a Find the one that cuts.

I 21 Find the picture that shows from what paper is made.

Similarities

S 1 Find the one that is not like the others.

Size and Number

N 3b Find the nest with the most eggs in it.

Immediate Recall

R 1 Find on this big card the picture that I just showed you.

The four orders of item placement gave basically four different forms of the PTI. These were: the published order, the absolute order, the random order, and the patterned order. A description of each form follows.

The published order of item arrangement was exactly as the test was published. The only change made was that special answer sheets were prepared for this study[1] and the items were numbered consecutively from 1 to 160. While the order of items in the other forms were changed, the assigned number from the published order was retained. A copy of the answer sheet used for the published order may be found in Appendix A along with the answer sheets for the other forms.

The absolute order of item arrangement consists of the items in ascending order of difficulty within each subtest.

---

[1] This study began before the PTI answer sheets were published.

The difficulty of each item was determined from the data received in the standardization of the PTI. Each item was subjected to an item analysis conducted on the responses made by all the subjects who attempted the item during the standardization. The arrangement of items in the absolute order is approximately the same as the published order but some differences do occur. These may be noted by referring to the answer sheet for this order which may be found in Appendix A.

The patterned order was composed by placing the items without regard to subtest. The two easiest items in the entire test were placed at the beginning of this form. They were followed by the hardest item. These three were followed by the next two easiest items and the next hardest item. This pattern of two easy items followed by one hard item was continued until all items were assigned positions. The level of difficulty of the easy items increased throughout the test while that of the hard items decreased. The difficulty of one easy and the hard item in the last sequence of three items was exactly the same. The one remaining item had the same difficulty level as the last two items in the previous sequence. In Appendix B the difficulty level of the items and the arrangement of the items in each form can be found.

It may be noted that there are only 137 response cards in the published test and 160 scorable responses. This difference occurs because some of the response cards are used more than once in a sequence of items. When the arrangement of items necessitated that this sequence of items be broken the response cards needed were photographically reproduced and placed in the proper position designated by that particular arrangement. Thus 160 response cards and items were used in all forms.

The second instrument used in this study was the _PTI Behavioral Rating Scale_. This scale was developed for use with the PTI by the test author and underwent several revisions before this study was undertaken. Although it was not designed specifically as an anxiety scale, the descriptions of behavioral patterns are some of those which may in general be used in a clinical description of behavior associated with anxiety within the testing situation. This scale differs from other psychometric instruments in that it provides a rating of observations during the test rather than before or after.

The _PTI Behavioral Rating Scale_ is based on observed differences between two behavioral descriptions placed on either end of a five point rating continuum as used in the semantic differential approach. Each item is scored one through five from the term on one side to the term on the other and the total rating score is obtained by summing the score for each item. This numbering procedure was established in order to eliminate zero or negative values for a total score.

The scale consists of 10 items which are basically observations of behavior and 14 items which are interpretations of the subject's reaction to the task and to the examiner. A copy of the entire PTI Behavioral Rating Scale may be found in Appendix C but four sample items appear below:

Must be told to look at cards	_ _ _ _ _	Studies cards carefully
Attention fixed on each card	_ _ _ _ _	Attention must be directed to many cards
Serious	_ _ _ _ _	Playful
Indifferent	_ _ _ _ _	Enthusiastic

The apparatus used in this study to record physiological behavior in terms of skin resistance was a Lafayette Multi-Purpose Polygraphy Recorder. The recorder has three paper drive speeds. The slowest drive speed, which was three inches per minute, was used in this study. The recorder was equipped with an oscillograph movement and an inking pen. An electrically operated impulser was also mounted in the recorder drive case.

The GSR amplifier is a balanced direct current amplifier with optional alternating current input coupling. Subject resistance was measured in a bridge circut, the subject appearing as one arm of the bridge and a calibrated potentiometer in the opposite arm.

In order to assure comparable readings in terms of the number of ohms of resistance per unit of pen deflection, the sensitivity potentiometer was calibrated at each five kilohms setting on the calibrated potentiometer or heliopot control. From this a regression equasion was obtained and the constant was added before each reading was made from the chart.

Calibration accuracy is a function of bridge source battery voltage. Before the study began it was found necessary to add an external voltmeter for checking the condition of the battery. If more than a five percent change was noted the battery was replaced. To eliminate another possible source of variance, the tubes in the amplifier were checked periodically and replaced when a loss of sensitivity was noted.

Finger electrodes were designed by the experimenter in order to assure constant contact with adjacent surfaces of the index finger of the subject's less preferred hand. The electrodes were zink disks one half inch in diameter and one sixteenth of an inch thick. The electrodes were attached to flexible cartridge wire which was connected to the amplifier leads. The electrodes were then mounted on circular sponge rubber pads of the same diameter as the electrodes which were then mounted on a sponge rubber strip one half inch wide, one half inch thick, and one inch long. This assembly was attached to a flexible plastic strip one-half inch wide and one inch long.

In order to attach the electrodes securely and to provide as much comfort as possible while doing so, a cloth fastner which has the trade name of "Velcro" was used. The fastner has two sections. Each section has a relatively smooth backing and one of the sections has a fuzzed surface while the other has a surface consisting of minute plastic hooks. When the hooked surface of the one section is placed on the fuzzed surface of the other section, the hooks become imbedded within the fuzzed material and provide a suprisingly secure bond.

The width of fastner which was used in this electrode construction was one inch. The backing of the two sections were placed together in such a manner so that a one inch strip of the hooked section extended beyond the end of the fuzzed section. The two sections were then sewn together on a sewing machine.

The plastic strip of the electrode assembly was glued to the smooth backing of the hooked section. This placed the fuzzed section next to the skin and held the electrodes against the palmar surface of the index finger.

The length of the fastner was sufficient to allow one and one half turns around the subject's finger. (See figure 1) This secured the electrode surface in such a manner that the hand could be moved without any apparent loss of contact. To assure maximum surface contact, the electrodes were curved to conform to the curvature of the finger.

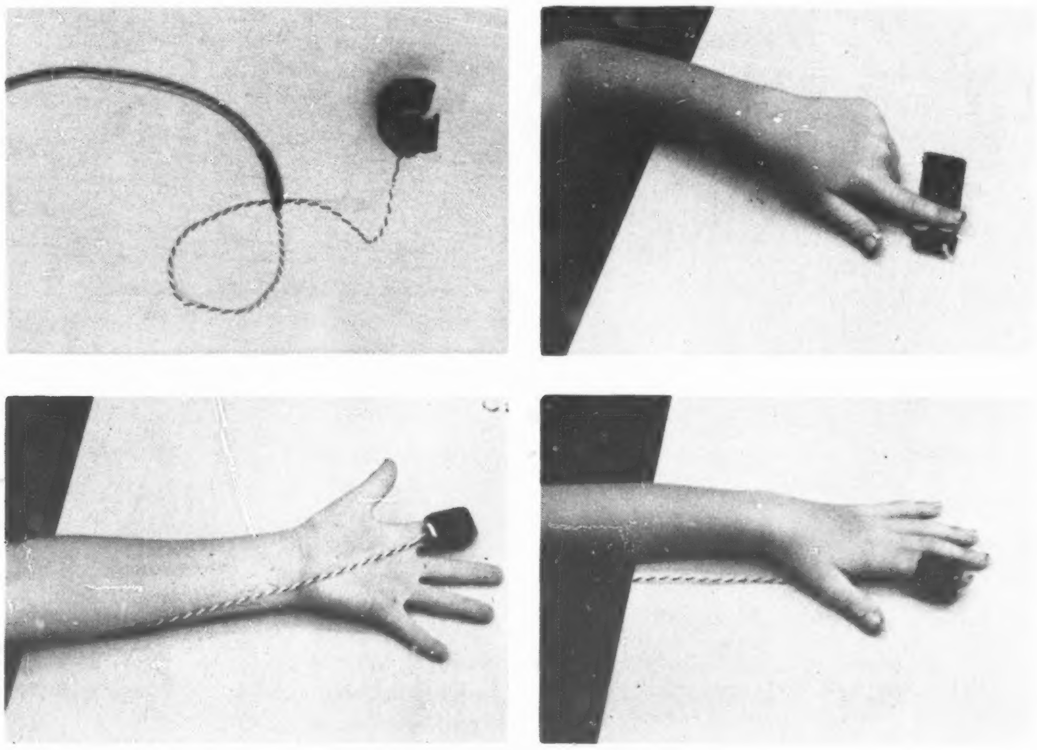

Fig. 1 Application of electrodes

To further insure a good contact, a zink sulfate electrode paste was made following the formula described by Lykken (1959, p. 634). The electrodes were cleaned after each use with a fine grain sandpaper to remove the oxidation and a thin coat of electrode paste was reapplied.

The subject's finger was cleaned before and after the attachment of the electrodes with alcohol to assure a clean skin surface. The procedure in the use of the above described electrodes and electrode paste was followed to reduce the measurement error as much as possible.

Method

The subjects were divided into four treatment groups as described previously. Each subject received each form of the PTI in a counter rotated order of administration. Subjects in Group I received the forms in the following order: published, random, absolute, and patterned. Group II subjects were administered the random order first followed by the absolute, patterned, and the published order. The order of administration to Group III was absolute, patterned, published, and random. Group IV received the patterned order first followed by the published, random, and absolute orders. This method of administration should remove any effects of sequence of testing with a particular form since no form was administered to more than one group of subjects first. A minimum of three and a maximum of six weeks elapsed between administrations of different PTI forms. The majority of subjects did not seem to be aware that the forms contained the same items. A few subjects stated that one or two items were like an item that was given before.

Since some of the response cards are the same for different items within the same test it is very likely that the references made were to cards previously used in the same form.

The data collected for each subject for each administration were: responses made to each item on the PTI form, observational data on the PTI Behavioral Rating Scale, and continuous skin resistance measured on the GSR recorder. The response latency was marked on the GSR strip chart by the use of the electrically operated impulser. A foot switch was added to allow the examiner to mark the period of time taken by the subject to respond after a response was requested.

All forms of the test were administered to all subjects and all of the rating scales were filled out by the same examiner. All tests were administered in testing rooms provided in the schools in which the subjects were enrolled. The conditions of light, temperature, and noise were very similar to those conditions which are normally found in these schools between the months of October and May and were judged by the examiner to be comfortable enough for optimum effort. All subjects were familiar with the rooms and the GSR recording apparatus since they had been exposed to the room and apparatus during the pretrial session mentioned eariler. The subjects were seated at a low table facing the examiner and the electrodes were attached to the index finger of the less perferred hand. The GSR recorder was placed behind the subject and was operated by a technician. The examiner engaged the subject in conversation by asking questions pertaining to the general school setting. When the technician indicated to the examiner that the input from the subject had been balanced on the GSR recorder, the test was introduced with the following statement:

> "(Subject's name) today I am going to give you a test to find out how much you know. I want you to try real hard and see if you can get all the answers right."

On the second, third and fourth administrations the instructions were as follows:

> "(Subject's name) today I am going to give you another test to see how well you can do. I want you to try real hard and see if you can get all the answers right."

The examiner remained passive through the entire length of the administration. Any statements made or questions asked by the subjects were merely reflected by the examiner. The subject was given no knowledge as to the adequacy of his performance either during or after the examination. When the subject asked if he were doing well, the question was reflected by the examiner. As the subjects were returned to the classroom after each of the first three tests they were informed that they would be given another test later. At the end of the last test, all subjects were praised for the hard work that they exhibited and informed that they had performed well on all the tests.

At the end of each test, the rating scales were filled out for each subject by the examiner and the appropriate entries were made on the GSR charts by the technician.

Analysis

At the conclusion of the testing program, raw data from three sources were available. These data were: responses made to the PTI items, positions assigned each item in the rating scale, and continuously recorded skin resistance levels. For the purpose of the analysis certain conversions of the raw data were necessary. Each of the methods of conversion will be discussed separately.

The responses made to the items on the various forms of the PTI were scored as being either correct or incorrect. The number of correct responses was summed to give a total raw score. The chronological age (CA) was computed for each subject at the time of testing. Using the total raw score and CA, a deviation IQ was obtained for each subject on each form from the tables constructed from the PTI standardization data.

Each item of the rating scale was given a numerical value ranging from one through five in keeping with the position checked by the examiner between the two descriptive statements. The direction of the scoring was based on an item analysis and only those items correlating significantly with the total score were retained for this study. These raw score values were summed to yield the total score and the total score was recorded for each subject under each condition.

The skin resistance level was recorded continuously from the beginning of the administration in terms of ohms of resistance. With continuous data recorded over a 30 to 45 minute period it soon became evident that a sampling technique would have to be employed. It was found satisfactory to record periodically the skin resistance level as the average of two consecuative items beginning with the first two items, skipping six items between readings and ending with the last two items in the test. This gave a total of 21 readings covering 42 of the 160 items in the test. The items being administered during this sampling are indicated with an asterisk on the item analysis sheets found in Appendix B. The GSR readings were recorded in terms of absolute ohms of resistance. The method of analyzing these scores will follow the statement of hypotheses.

The general hypothesis was that no differences will be found in intelligence scores as measured by a point scale, psychological behavior as observed by an examiner, or in physiological reaction as measured by a galvanometer. The specific hypotheses tested were:

there will be no differences between the mean PTI IQs attributable to the administration of the four different orders of item arrangement.

there will be no differences between the mean PTI IQs attributable to the four times the PTI was administered.

there will be no differences between the mean PTI Behavioral Rating Scale scores attributable to the administration of the four different orders of item arrangement.

there will be no differences between the mean PTI Behavioral Rating Scale scores attributable to the four times the PTI was administered.

there will be no differences between means of the GSR skin resistance measures attributable to the administration of the four different orders of item arrangement.

there will be no differences between the means of the GSR skin resistance measures attributable to the four times the PTI was administered.

It was further hypothesized that:

there will be no significant relationship between each measure and each other measure.

In order to test these hypotheses, the appropriate statistical techniques were selected. The counter rotated order of administration provided a greco-latin square design with a four x four matrix. The order of items for each administration provided the four column designations and the four times the PTI was administered provided the row designations. Thus a treatment x time matrix was established. The appropriate statistical technique for the analysis of mean scores with this design is the analysis of variance technique.

In order to test the hypothesis of no relationship between measures a Pearson product-moment coeficient of correlation procedure was selected.

One problem became apparent when the analysis of variance technique was selected. The data from the GSR skin resistance measure could not be compared directly since the readings obtained were resistance levels in terms of absolute ohms of resistance. If the readings were summed and a mean score was obtained it would be a function of the subject's initial resistance level. In most measurement procedures the variables are generally measured from a zero point. This is not the case with resistance measures since the subject had some skin resistance before the experimental treatment. The initial level of resistance not only varied between subjects but for the same subject from time to time.

This being the case, it was necessary to remove the effects of the intital level of resistance on the single index of the GSR.

The GSR data were converted to a mean change score by subtracting each of the sampled readings from the initial level exhibited by the subject. The resulting values were assigned a numerical value either plus or minus depending on whether the value of the reading was above or below the initial level. These values were then summed algrebraically and a mean change score was determined for each subject. This value was entered as the single index of GSR for the analysis of variance problem.

The above procedure provides a single index of skin resistance for each subject under each condition of testing. An index so derived may conceivably have one major defect. This defect arrises from the representativeness of any single index. With a single index, the question which is inherent within its use is whether such an index can adequately represent the total descriptiveness of the individual values which comprise that index. This problem is common to all single indices which may be used to represent a given class of behaviors. The assumption of the representativeness of the mean change score is as sound as the assumption that an IQ is representative of the total behavior of a an individual when performing the tasks required on an individual intelligence test. Therefore, the use of a mean change score to represent physiological reactions was not considered to be a deterent to the analysis used in this study.

In this chapter the subjects were described and the instruments and apparatus used were described. The method of treatment and the data collection procedures were discussed. The techniques involved in the statistical analysis were set forth. In the following chapter the results will be presented and will be briefly discussed.

Chapter 4

Results and Discussion

The results of the analysis of the data are presented in this chapter as they relate to each of the hypotheses set forth in the previous chapter. The general hypothesis stated earlier was that there will be no differences found in intelligence test scores as measured with a point scale, in psychological behavior as observed by an examiner, or in physiological reaction as determined with a galvanometer. It was further hypothesized that there would be no significant relationship between the measures. The results will be presented in connection with each of the specific hypotheses.

The first specific hypothesis for which data were submitted for statistical analysis was that there will be no significant differences between mean PTI IQs attributable to the administration of the four different orders of item arrangement. This hypothesis was not rejected. The results of the analysis of variance procedure used to test this hypothesis are found in Table 3.

The second specific hypothesis was that there will be no differences between the mean PTI IQs attributable to the four times the PTI was administered. This hypothesis was rejected. The results may also be found in Table 3 under the designation of Time.

Table 3

Summary of
Analysis of Variance
for PTI IQ means

Source	DF	MS	F	
Time	3	1101.5328	14.131	**
Order	3	89.3881	1.146	
Interaction TXO	9	46.5752	1.673	
Residual	592	77.9469		
Total	607	82.5972		

** Significant at .01 level

The differences between the mean IQs were significant for the time of testing only. The order of item presentation did not yield significant differences in the mean scores. In order to further determine the magnitude of these differences obtained according to rows (time that the PTI was administered) the means for each row were computed and t tests were performed. The t tests for each pair of means indicated that there were significant differences between the mean of the first time of testing and all other times and between the mean of the second and fourth times of testing. Neither difference between the means of the second and the third times of testing nor the difference between the means of the third and the fourth times of testing were significant. The means, mean differences and t values for the time the PTI was administered may be found in Table 4. The means, mean differences and t values for the administration of the four different orders of item arrangement may be found in Table 5.

Table 4

Means, Mean Differences and
t values for time of
Administration

Means				Diff.	t
First	112.7763 -	Second	116.8289	4.0526	3.7602 **
First	112.7763 -	Third	117.6578	4.8815	4.4535 **
First	112.7763 -	Fourth	119.0394	6.2631	5.6736 **
Second	116.8289 -	Third	117.6578	.8289	.8476
Second	116.8289 -	Fourth	119.0394	2.2105	2.2100 *
Third	117.6578 -	Fourth	119.0394	1.3816	1.3611

** Significant at .01 level

* Significant at .05 level

Table 5

Means, Mean Differences and t values for Order of Item Presentation

Means				Diff.	t
Pub.	116.0657	- Ran.	117.7039	1.6382	1.5236
Pub.	116.0657	- Abs.	116.1250	.0593	.0553
Pub.	116.0657	- Pat.	116.4078	.3421	.3277
Ran.	117.7039	- Abs.	116.1250	1.5789	1.5031
Ran.	117.7039	- Pat.	116.4078	1.2961	1.2759
Abs.	116.1250	- Pat.	116.4078	.2828	.2741

The two hypotheses of no significant differences between the mean PTI Behavioral Rating Scale scores were not rejected. The analysis of variance data indicated only one significant F value at the .05 level of confidence and that was the F value for the interaction between the order of item presentation and the time in the administrational sequence. A summary of the analysis of variance data may be found in Table 6. The means for the order of item arrangement and for the times of testing may be found in Table 7.

Table 6

Summary of Analysis of Variance for mean PTI rating scale scores

Source	DF	MS	F
Time	3	25.9758	6.330
Order	3	196.0592	1.192
Interaction TXO	9	57.7302	2.848 *
Residual	592	164.4428	
Total	607	162.3325	

* Significant at .05 level

Table 7

Mean PTI Behavioral Rating Scale scores for Order of Item Arrangement and Time of Testing

Order		Time	
Published	66.7565	First	65.5855
Random	65.1250	Second	65.7236
Absolute	67.1513	Third	66.1250
Patterned	64.9013	Fourth	66.5000

The two hypotheses of no difference between the means of the GSR resistance measures were not rejected. The analysis of variance data for the GSR measures may be found in Table 8. The means according to order of item arrangement and time of testing may be found in Table 9.

Table 8

Summary of
Analysis of Variance
for GSR means

Source	DF	MS	F
Time	3	2858480930	2.010
Order	3	1023469970	4.014
Interaction TXO	9	8056810583	1.960
Residual	592	4108744867	
Total	607	4172543837	

Table 9

Mean GSR Change Scores for Order
of Item Arrangement and Time of Testing

Order		Time	
Published	-55,462.3356	First	-48,979.6185
Random	-58,506.9869	Second	-51,874.8553
Absolute	-53,605.5329	Third	-65,040.0922
Patterned	-58,645.9343	Fourth	-59,981.2237

The analysis of variance procedure gave no F value large enough to be significant at either the .05 or the .01 level.

The hypothesis of no relationship between the measures was rejected. The correlations between the IQs and rating scale scores were significantly different from zero and in the negative direction. The correlations between the two measures may be found in Table 10.

Table 10

Correlation Matrix
between IQs and rating
scale scores

	1st	2nd	3rd	4th	Total
Published	-.69	-.37	-.60	-.50	-.55
Random	-.07*	-.63	-.16*	-.46	-.30
Absolute	-.42	0.43	-.62	-.58	-.49
Patterned	-.45	-.36	-.52	-.72	-.51
Total	-.42	-.47	-.50	-.56	-.46

* Values not significant at .05 level

From the correlational values found in Table 10 it may be seen that all of the values are in the negative direction. All but two of these values are significant; those which are not are marked with an asterick.

The relationship between the mean GSR measures and the two other measures were not significantly different from zero. The correlation between the GSR measure and PTI IQs was .019 and the correlation between the GSR measures and Rating Scale scores was -.026.

The results of the analysis of mean PTI IQ differences within this study suggest several explanations. The increase in mean scores for the four times the tests were administered may be attributed to the "practice effect" which is sometimes expected when a test is administered twice within a three to six week period. However, differences of this magnitude have not been found with smaller samples. The difference between the first administration and the second was the largest being approximately four points. The gain in mean score between the second and third, and the third and fourth administration were somewhat smaller and the one and two point gain, respectively, may indicate that "practice effect" contributes the major portion of gain in the second testing period when that period follows the first testing period by a few weeks. Also, it is possible that the gain in raw score is attributed to information provided during regular classroom activities. An analysis of the specific items contributing to the increased score has not been attempted. Further, it could be hypothesized that the children were not as "ready" for the first administration as those which followed. Although they responded to the TASC and GASC while the galvanometer was operating, such activity may not have been sufficient. An analysis of GSR basal levels was not included in this study but these data should receive further attention.

The finding of no significant differences between the mean PTI IQs attributable to the four orders of item arrangement could be explained in several ways. One suggests that if each item tends to measure that portion of behavior for which it was designed and is not dependent upon the other items within the test, the sequence in which the items are placed should not effect the total score. A heavy loading with a single common factor could help explain the relationship of each item to the total score.

The norms derived from the standardization procedure apparently are valid for groups of subjects even though the items are administered in an order different from the published order. This generalization must be made with caution, however. The arrangement of items may not be a factor for groups of subjects, but may very well be a factor in the performance of some subjects within those groups or for subjects appreciably different from those used in this study.

The differences between the mean Rating Scale scores were obviously not significant for either time of testing or for the order of item arrangement. The slight, but nonsignificant, increase in the mean scores for time of testing may well have been the result of the "practice effect" on the part of the examiner in that increasing skill was acquired as the number of observations increased.

The slight, but nonsignificant, differences between the mean Rating Scale scores for the order of item arrangement may very well be due to chance, but the pattern of these differences is somewhat more difficult to explain on a chance basis alone. The means for the Absolute order is the greatest and this is followed by the Published order mean. Both these means are higher than the mean for the Random order which in turn is higher than the mean for the Patterned order. If the arrangement of items is not a factor for groups of subjects, the trend in these mean scores may be explained on the basis that the order of item arrangement is a factor in the performance of a small minority of subjects within each group. Also, the arrangement of items may present a sequence which influences observable reaction to the individual tasks but does not effect the solution of these tasks. This, of course, would require further analysis to determine if this were the case.

The correlation between PTI IQs and Rating Scale scores was significantly different from zero and reached moderate proportions for both time of testing and for order of item arrangement. The degree of this relationship is not unexpected since the correlation is based on observed behavior and performance in the same setting.

From the negative relationship it can be seen that there is a tendency for those subjects who received high scores on the Rating Scale to receive low PTI IQs. Since on the Rating Scale a high score indicates inefficient test behavior, and since the descriptions of behavior on the scale are some of those which could generally be considered descriptive of anxious behavior, it would seem that a relationship between observed behavior and performance on the PTI exists. Since there were no differences between group means on either the PTI or Rating Scale and since a moderate relationship between observed behavior and performance exists, it would seem that behaviors observed is more of a factor for some individuals than for others.

The findings in this study seem to support the contention that test item arrangement will not be a critical factor in the behavior of similar groups of first grade children responding to similar tasks but may be a factor for some individuals within the groups.

Superficially it would be expected that the arrangement of items within a test would effect examinee performance. The results of the statistical analysis indicate that this is not the case since the mean IQs derived from the four different orders of item arrangement were not significantly different. This finding seems to support the position that each item, within the point scale used in this study, is independent of the other items within the test except as they each measure one common general factor. Each item measures that portion of behavior for which it was designed without support from other items of the same type. The location of difficult items did not effect the over-all performance of the groups of children in this study. It may be noted that while the mean scores for the Random order and the Patterned order were higher than those for the Published and Absolute orders, the differences were so small that they must be explained as chance differences. For comparable groups of children it is not likely that the manipulation of items within a similar test would provide an advantage or disadvantage.

The high relationship found between observed behavior and performance on the PTI would seem to open another avenue for consideration. If the arrangement of items does not appreciable effect examinee performance and if certain observed behaviors tend to be inversely related to superior performance, then a manipulation of items within the test to place some subjects more at ease may be carried out without adversely effecting the validity of the score.

In this chapter the results of the statistical analyses were presented and the implications of these results were discussed. In the following chapter, the summary and the conclusions will be presented along with recommendations for further research.

Chapter 5

Summary, Conclusions and Recommendations

Summary

This study was undertaken as a result of a growing concern on the part of many test users about the occasional failure of tests to adequately assess potential behavior of some examinees. This occasional failure may be due to some factors which are inherent in the testing situation. In any testing situation there are three basic elements: the examiner, the examinee and the test. In order to maximize examinee performance certain suggestions have been made which involve one of these elements -- the examiner and his behavior. The success of the examiner in altering his behavior to maximize examinee performance depends entirely on the skill of the examiner and his awareness. In trying to minimize the importance of examiner skill and awareness, the manipulation of the test became the focal point of this study.

The purpose of this investigation was to study the effects of test item arrangement on the psychological and physiological behavior of primary-school children. Specifically, the study was undertaken to determine the effects of four different arrangements of items within a new point scale on the subject's performance on the intelligence measure, on psychological behavior as measured with a galvanometer.

In the literature reviewed, it was established that anxiety in the testing situation arises out of the evaluative nature of the testing procedure, that anxiety has been measured with both physiological and psychological indices and that anxiety is a factor in performance. Anxiety seems to be significantly related to the quality, efficiency and type of verbal performance. The level of anxiety seems to be inversely related to academic performance, seems to effect speed of reaction, and have an adverse effect on concept formation and concept judgement. In the area of performance on intelligence tests, anxiety seems to have a dehabilitating effect which may be mediated by the format of the test and the removal of time limits.

The design of the study involved selecting 152 pupils who had earned CTMM IQs between 100 and 125 and who were enrolled in a first grade. Each subject selected for inclusion in the study was administered the _General Anxiety Scale for Children_ and the _Test Anxiety Scale for Children_ verbally in a pre-experimental setting while a GSR recorder was attached and in operation. An equal number of boys and girls were assigned randomly to one of four groups whose mean CTMM IQs or mean TASC scores were not significantly different from those means for the other groups.

The instruments used in this study were the PTI and the PTI Behavioral Rating Scale. The PTI is a point scale designed to measure general intelligence and yields a deviation IQ. The PTI is individually administered and contains 160 items arranged in six subtests. The subject may respond by selecting the response picture he considers to be appropriate to the examiner's request. Little physical activity is required.

The PTI Behavioral Rating Scale contains 10 items which are descriptive of subject behavior and 14 items which are descriptive of the subject's reaction to the task or to the examiner. The items are composed of statements placed at either end of a five point rating continuum following the semantic differential approach. The numerical value assigned to each of the positions to be checked was summed to yield a total score.

The apparatus used to measure skin resistance was a Lafayette Multipurpose Polygraph Recorder which was equipped with a GSR amplifier and movement, an incident marker, and a three speed chart drive motor. The speed of three inches per minute was used in this study.

To facilitate accurate measurement, finger electrodes were specifically designed and constructed for this study and calibrations were made for the instrument.

The items of the PTI were arranged in four different orders: Published, Absolute, Random and Patterned. In the Published order the items were arranged in the estimated order of difficulty within each subtest prior to the standardization while in the Absolute order the items were slightly rearranged according to the standardization data. In the Random order the items were assigned positions throughout the test without regard for the subtests. In the Patterned order the items were arranged without regard for subtests and in a sequence of two easy items followed by one difficult item ranging from the easiest and hardest at the beginning of the test to three items of equal difficulty at the end of the test.

The four arrangements of PTI items were administered individually to the subjects in each group in a counter-rotated order while continuous skin resistance measures were recorded. The Rating Scale was filled out for each subject under each condition of testing.

Data for each of the four groups of subjects from three sources (IQ, Rating Scale scores, GSR measures) was then available. This data was submitted to statistical analysis to aid in the testing of the null hypotheses which were:

1. there will be no differences between the mean PTI IQs which can be attributed to the administration of the four orders of item arrangement.

2. there will be no differences between the mean PTI IQ which can be attributed to the time of testing.

3. there will be no differences between the mean Rating Scale scores which can be attributed to the administration of the four orders of item arrangement.

4. there will be no differences between the mean Rating Scale scores which can be attributed to the time of testing.

5. there will be no difference between the mean skin resistance measure which can be attributed to the administration of the four orders of item arrangement.

6. there will be no differences between the mean skin resistance measures which can be attributed to the four times of testing.

7. there will be no significant relationships between the measures.

The hypotheses dealing with differences between the mean PTI IQs attributable to the four times of testing and the relationship between the measures were rejected. The other hypotheses were not.

Conclusions

The conclusions which may be drawn from the work with the subjects of this study are:

1. the arrangement of items within the PTI does not effect the performance of first grade subjects in terms of IQ.

2. with repeated administrations of the PTI there was an increase in performance level of the subjects.

3. the order of PTI item arrangement in itself is not a factor which leads to observable differences in overt behavior.

4. the order of PTI item arrangement or the number of times of testing does not unequally effect anxiety as measured by amount of change in skin resistance.

5. a significant relationship exists between superior performance on the PTI and observed behaviors which are commonly associated with appropriate awareness in the testing situation.

Recommendations

Although the null hypotheses formulated for this study were infrequently refuted, the authors are still concerned about the effects of stress generated by items in tests of intelligence. Perhaps significant findings would have resulted if subjects who were known to be anxious in testing situations were used in the study. Another possibility is that significant results may have resulted if a wider range of intelligence had been sampled instead of the range between 100 and 125. Perhaps a pilot study with a few children who are recommended by teachers as "test anxious" may be the next step.

The age of the subjects may have been a factor also. It is possible that young children are not as anxious as older children who realize that individual testing may be an evaluative procedure in the school program. It would be interesting to study both older children and children of the same age as were used in this study to see if item arrangement has differential effects on the arousal level of the two groups. Also, it would be interesting to study the effects of items which are more arousing than those used in this study.

Since PTI IQs in this study increased with repeated test administrations and repeated exposure of the subjects to the examiner, it would be enlightening to investigate the relationship of the scores earned by similar groups of children when the test is administered by a person who is well known to them, such as their teacher, and by an examiner who frequently visits the classroom and is well known, and by an examiner who is not well known. Other physiological measures may be used in connection with the studies which have been suggested. A total pattern of physiological reaction may show differences in arousal where a single index may fail to do so.

References

Alpert, R. & Haber, R. N. Anxiety in academic achievement situations. J. abnorm. soc. Psychol., 1960, 61, 207-215.

Atkinson, J. W. & Litwin, G. H. Achievement motive and test anxiety conceived as motive to approach success and motive to avoid failure. J. abnorm. soc. Psychol., 1960, 60, 52-63.

Barnard, J. W., Zimbardo, P. G. & Sarason, S. B. Anxiety and verbal behavior in children. Child Develpm., 1961, 32, 379-392.

Berry, J. L. & Martin, B. GSR reactivity as a function of anxiety, instructions, and sex. J. abnorm. soc. Psychol., 1957, 54, 9-12.

Bitterman, M. E. & Holtzman, W. H. Conditioning and extenction of the galvanic skin response as a function of anxiety. J. abnorm. soc. Psychol., 1952, 47, 615-623.

Bixenstein, V. E. A case study of the use of palmar sweating as a measure of psychological tension. J. abnorm. soc. Psychol., 1955, 50, 138-141.

Buchwald, A. M. Manifest anxiety-level, verbal response-strength, and paired-associate learning. Amer. J. Psychol., 1959, 72, 89-93.

Buskirk, C. V. Performance on complex reasoning tasks as a function of anxiety. J. abnorm. soc. Psychol., 1961, 62, 201-209.

Callahan, R. J. & Keller, J. E. Digit span and anxiety: An experimental group revisited. Amer. J. ment. Defic., 1957, 61, 581-582.

Calvin, A. D., McGuigan, F. J., Tyrell, Sybil & Soyars, M. Manifest anxiety and the palmar perspiration index. J. consult. Psychol., 1956, 20, 356.

Castaneda, A. Reaction time and response amplitude as a function of anxiety and stimulus intensity. J. abnorm. soc. Psychol., 1956, 53, 225-228.

Castaneda, A., Palermo, D. S. & McCandless, B. R. Complex learning and performance as a function of anxiety in children and task difficulty. *Child. Develpm.*, 1956, *27*, 327-332.

Davidson, W. Z., Andrews, T. G. & Ross, S. Effects of stress and anxiety on continuous high-speed color naming. *J. exp. Psychol.*, 1956, *52*, 13-17.

Du Tiot, J. M. Skin conductance and peripheral vascular reactions. *J. exp. Psychol.*, 1956, *52*, 392-397.

Feldhusen, J. F. & Klausmeier, H. J. Anxiety, intelligence, and achievement in children of low, average and high intelligence. *Child Develpm.*, 1962, *33*(2), 403-409.

French, J. L. *Pictorial Test of Intelligence*. Boston: Houghton Mifflin, 1964.

Gilberstadt, H. & Davenport, G. Some relationships between GSR conditioning and judgments of anxiety. *J. abnorm. soc. Psychol.*, 1960, *60*, 441-443.

Goodstein, L. D. & Farber, I. E. On the relationship between A - scale scores and digit symbol perference. *J. consult. Psychol.*, 1957, *21*, 152-154.

Gordon, J. E., Martin, B. & Lundy, R. M. GSRs during repression, suppression, and verbalization in psychotherapeutic interviews. *J. consult. Psychol.*, 1959, *23*, 243-251.

Grooms, R. R. & Endler, N. S. The effect of anxiety on academic achievement. *J. educ. Psychol.*, 1960, *51*, 299-304.

Gynther, Ruth A. The effects of anxiety and of situational stress on communicative efficiency. *J. abnorm. soc. Psychol.*, 1957, *54*, 274-276.

Hafner, A. J., Pollie, D. M. & Wapner, I. The relationship between the CMAS and WISC functioning. *J. Clin. Psychol.*, 1960, *16*, 322-323.

Keller, E. D. & Rowley, V. N. Anxiety, intelligence and scholastic achievement in elementary school children. *Psychol. Rep.*, 1962, *11*, 19-22.

Kerrick, Jean S. The effects of manifest anxiety and IQ on discrimination. J. abnorm. soc. Psychol., 1956, 52, 136-138.

Kimmel, H. D. & Hill, Frances A. A comparison of two electrodermal measures of response to stress. J. comp. physiol. Psychol., 1961, 54, 395-397.

L'Abate, Luciano. Manifest anxiety and the learning of syllables with different associative values. Amer. J. Psychol., 1959, 72, 107-110.

Lawrence, S. W., Jr. The effects of anxiety, achievement motivation, and task importance upon performance on an intelligence test. J. educ. Psychol., 1962, 53(3), 150-156.

Learmonth, G. J., Ackerly, W. & Kaplin, M. Relationship between palmar skin potential during stress and personality variables. Psychosom. Med., 1959, 21, 150-157.

Leavitt, H. C. Genetic relations between anxiety and fear. Psychoanal. Rev., 1957, 44, 65-72.

Lighthall, F., Ruebush, B., Sarason, S. B. & Zweibelson, I. Change in mental ability as a function of test anxiety and type of mental test. J. Consult. Psychol., 1959, 23, 34-38.

Lipman, R. S. & Griffith, B. C. Effects of anxiety level on concept formation: A test of drive theory. Amer. J. ment. Defic., 1960, 65, 342-348.

Lotsof, E. J. & Downing, W. L. Two measures of anxiety. J. consult. Psychol., 1956, 20, 170.

Lykken, D. T. Properties of electrodes used in electrodermal measurement. J. comp. Physiol. Psychol., 1959, 52, 629-634.

McCandless, B. R. & Castanada, A. Anxiety in children, school achievement, and intelligence. Child Develpm., 1956, 27, 379-382.

McDonnell, G. J. & Carpenter, J. A. Anxiety, skin conductance and alcohol: A study of the relation between anxiety and skin conductance and the effect of alcohol on the conductance of subjects in a group. Quart. J. Stud. Alcohol, 1959, 20, 38-52.

McDonnell, G. J. & Carpenter, J. A. Manifest anxiety and prestimulus conductance levels. J. abnorm. soc. Psychol., 1960, 60, 437-438.

Mandler, G., Mandler, Jean M. & Uviller, Ellen T. Autonomic feedback: The perception of autonomic activity. J. abnorm. soc. Psychol., 1958, 56, 367-373.

Martin, B. The validity of a self report measure of anxiety as a function of the time interval covered by the instructions. J. consult. Psychol., 1959, 23, 468.

Martin, B. The assessment of anxiety by physiological behavioral measures. Psychol. Bull., 1961, 58, 234-255.

Merwin, J. C. & Gardner, E. F. Development and application of tests of educational achievement. Rev. educ. Res., 1962, 32, 40-50.

Novak, S., Moriarty, D., Koltuy, M. & Welch, L. The effect of increased trials on the stability of the PGR in an avoidance situation. J. Psychol., 1956, 41, 243-246.

Phillips, B. N. Sex, social class, and anxiety as sources of variation in school achievement. J. educ. Psychol., 1962, 53, 316-322.

Pickrel, E. W. The differential effect of manifest anxiety on test performance. J. educ. Psychol., 1958, 49, 43-46.

Raphelson, A. C. The relationship among imaginative, direct verbal, and physiological measures of anxiety in an achievement situation. J. abnorm. soc. Psychol., 1957, 54, 13-18.

Raphelson, A. C. & Moulton, R. W. The relationship between imaginative and direct verbal measures of test anxiety under two conditions of uncertainty. J. Pers., 1958, 26, 556-567.

Rawn, M. L. The overt-covert anxiety index and hostility. J. clin. Psychol., 1958, 14, 279-280.

Ringness, T. A. GSR during learning activities of children of low, average, and high intelligence. Child Develpm., 1962, 33(4), 879-889.

Rosenstein, A. J. Psychometric versus physiological anxiety and serial learning. J. Pers., 1960, 28, 279-292.

Ruebush, B. K. Interfering and facilitating effects of test anxiety. J. abnorm. soc. Psychol., 1960, 60, 205-212.

Sarason, I. G. Effect of anxiety, motivational instructions and failure on serial learning. J. exp. Psychol., 1956, 51, 253-260.

Sarason, I. G. The effect of anxiety and two kinds of failure on serial learning. J. Pers., 1957a, 25, 383-392.

Sarason, I. G. Test anxiety, general anxiety, and intellectual performance. J. consult. Psychol., 1957b, 21, 485-490.

Sarason, I. G. Effects on verbal learning of anxiety, reassurance, and meaningfulness of material. J. exp. Psychol., 1958, 56, 472-477.

Sarason, I. G. A note on anxiety, instructions, and word association performance. J. abnorm. soc. Psychol., 1961, 62, 153-154.

Satason, S. B., Davidson, D. S., Lighthall, F. F., Waite, R. R., Ruebush, B. K. Anxiety in elementary school children. A report of research. New York: Wiley & Sons, 1960.

Shepard, A. H. & Abbey, D. S. Manifest anxiety, and performance on a complex perceptual-motor task. Percept. mot. Skills, 1958, 8, 327-330.

Siegman, A. W. The effect of manifest anxiety on a concept formation task, a nondirected learning task, and on timed and untimed intelligence tests. J. consult. Psychol., 1956, 20, 176-178.

Silverman, R. E. & Blitz, B. Learning and two kinds of anxiety. J. abnorm. soc. Psychol., 1956, 52, 301-303.

Sines, J. O. Conflict-related stimuli as elicitors of selected physiological responses. J. proj. Tech., 1957, 21, 194-198.

Sinha, D. & Singh, T. R. Manifest anxiety and performance on problem solving tasks. J. consult. Psychol., 1959, 23, 469.

Smock, C. D. The relationship between test anxiety, "threat-expectancy", and recognition thresholds for words. J. Pers., 1956, 25, 191-201.

Spence, K. W., Farber, I. E. & McFann, H. H. The relation of anxiety (drive) level to performance in competitional paired-associates learning. J. exp. Psychol., 1956, 52, 296-205.

Spielberger, C. D. & Katzenmeyer, W. G. Manifest anxiety, intelligence, and college grades. J. consult. Psychol., 1959, 23, 278.

Stanley-Jones, D. The structure of emotion: II. The physical basis of anxiety. J. nerv. ment. Dis., 1957, 125, 247-258.

Stevenson, H. W. & Iscoe, I. Anxiety and discriminative learning. Amer. J. Psychol., 1956, 69, 113-114.

Sullivan, M. & Calvin, A. D. Spoken and written vocabulary: Their relation to a standard vocabulary test, intelligence and anxiety. J. consult. Psychol., 1957, 21, 160.

Venables, P. H. Some findings on the relationship between GSR and motor task performance. J. gen. Psychol., 1956, 55, 199-202.

Waite, R. R., Sarason, S. B., Lighthall, F. F. & Davidson, K. S. A study of anxiety and learning in children. J. abnorm. soc. Psychol., 1958, 57, 267-270.

Whiteman, M. Anticipated frustration as a determinant of anxiety. J. gen. Psychol., 1957, 57, 183-195.

Wiener, G. The interaction among anxiety, stress instructions, and difficulty. J. consult. Psychol., 1959, 23, 324-328.

Zweibelson, I. Test anxiety and intelligence test performance. J. consult. Psychol., 1956, 20, 479-481.

APPENDIX A

Answer Sheets for the Pictorial Test of Intelligence

PUBLISHED ORDER

NO ______

IQ ______

Name ______________________ Date ________________ RS ______

School ____________________ B D ________________ CA ______

Picture Vocabulary

1.	Chair	TLBR R
2.	Stool	TLBR R
3.	Axe	TLBR T
4.	Farm	TLBR B
5.	Pitcher	TLBR L
6.	Horse	TLBR L
7.	Dessert	TLBR R
8.	Arm	TLBR L
9.	Sill	TLBR R
10.	Cord	TLBR L
11.	Tree	TLBR B
12.	Shrubbery	TLBR T
13.	Refreshment	TLBR B
14.	Weapon	TLBR L
15.	Finished	TLBR R
16.	Uniform	TLBR L
17.	Collision	TLBR T
18.	Autumn	TLBR B
19.	Padlock	TLBR B
20.	Anchor	TLBR L
21.	Nose	TLBR T
22.	Manufacturing	TLBR T
23.	Swift	TLBR L
24.	Ascend	TLBR L
25.	Clock	TLBR R
26.	Coat	TLBR R
27.	Retire	TLBR T
28.	Knot	TLBR T
29.	Plunge	TLBR L
30.	Bolt	TLBR R
31.	Astonishment	TLBR R
32.	Acquiring	TLBR T

Info-Compre

60.	Cuts	TLBR R
61.	Presses clothes	TLBR T
62.	Pour	TLBR L
63.	Temperature	TLBR L
64.	Keeps dry	TLBR T
65.	Used to fight	TLBR L
66.	Helps us see	TLBR B
67.	Helps us smell	TLBR T
68.	Grow ground	TLBR B
69.	People eat most	TLBR T
70.	Man rides	TLBR L
71.	I could carry	TLBR B
72.	G Y R B	+ -
73.	Butterfly	TLBR B
74.	Used W/needle	TLBR L
75.	Used W/broom	TLBR L
76.	Run fastest	TLBR L
77.	Milk people	TLBR R
78.	Heaviest	TLBR B
79.	Carry most	TLBR L
80.	Seal eats	TLBR L
81.	Real smallest	TLBR L
82.	Real biggest	TLBR T
83.	Man always	TLBR B
84.	Quarter past 8	TLBR T
85.	Paper made	TLBR R
86.	Not eyelid	TLBR R
87.	House must	TLBR R
88.	Closest if real	TLBR L

Form Discrimination

33.	TLBR B
34.	TLBR R
35.	TLBR R
36.	TLBR B
37.	TLBR T
38.	TLBR T
39.	TLBR R
40.	TLBR B
41.	TLBR T
42.	TLBR R
43.	TLBR L
44.	TLBR T
45.	TLBR L
46.	TLBR R
47.	TLBR R
48.	TLBR R
49.	TLBR T
50.	TLBR L
51.	TLBR B
52.	TLBR L
53.	TLBR T
54.	TLBR L
55.	TLBR L
56.	TLBR B
57.	TLBR L
58.	TLBR R
59.	TLBR T

PUB. O.

NO. _______

Similarities

89. TLBR L
90. TLBR B
91. TLBR T
92. TLBR T
93. TLBR L
94. TLBR T
95. TLBR T
96. TLBR L
97. TLBR R
98. TLBR L
99. TLBR L
100. TLBR T
101. TLBR R
102. TLBR B
103. TLBR T
104. TLBR T
105. TLBR R
106. TLBR B
107. TLBR B
108. TLBR R
109. TLBR B
110. TLBR T

Size and Number

111. Biggest + -
112. Smallest + -
113. Longer 3/3 5/6 + -
114. Most TLBR B
115. Four TLBR T
116. Three TLBR R
117. Most TLBR L
118. Six TLBR B
119. Eight TLBR L
120. 5 3 4 2 + -
121. Two TLBR T
122. Three TLBR L
123. Three TLBR L
124. Four TLBR B
125. 9 7 8 6 + -
126. Six TLBR T
127. Eight TLBR B
128. Eight TLBR B
129. 3 + 2 TLBR L
130. 4 + 3 TLBR T
131. 5 - 1 - 1 TLBR T
132. 12 - 5 TLBR B
133. 13 - 8 TLBR T
134. 4/12 TLBR L
135. Whole 1/2 TLBR R
136. Whole 1/4 TLBR L
137. Whole 2/3 TLBR B
138. 24 - 18 TLBR B
139. 12 + 6 TLBR T
140. 4 jacks 1/2 TLBR R
141. 35 + 10 (2) 5 TLBR L

Immediate Recall

142. TLBR R
143. TLBR T
144. TLBR T
145. TLBR R
146. TLBR L
147. TLBR T
148. TLBR R
149. TLBR R
150. TLBR L
151. TLBR L
152. TLBR B
153. TLBR L
154. TLBR B
155. TLBR L
156. TLBR R
157. TLBR B
158. TLBR R
159. TLBR R
160. TLBR T

NOTES

RANDOM ORDER

NO ______

IQ ______

Name ________________________ Date ____________________ RS ______

School ______________________ B D ____________________ CA ______

104.	Not like the	TLBR	T	91.	Not like the	TLBR	T
94.	Not like the	TLBR	T	133.	13 - 8	TLBR	T
103.	Not like the	TLBR	T	40.	Looks just like	TLBR	B
71.	I could carry	TLBR	B	16.	Find uniform	TLBR	L
23.	Find swift	TLBR	L	148.	Show picture on	TLBR	R
10.	Find cord	TLBR	L	105.	Not like the	TLBR	R
70.	Man rides	TLBR	L	139.	12 + 6	TLBR	T
24.	Find ascend	TLBR	L	66.	Helps us see	TLBR	B
7.	Find dessert	TLBR	R	2.	Find stool	TLBR	R
53.	Looks just like	TLBR	T	45.	Looks just like	TLBR	L
5.	Find pitcher	TLBR	L	130.	4 + 3	TLBR	T
97.	Not like the	TLBR	R	125.	9 7 8 6	+ -	
145.	Show picture on	TLBR	R	151.	Show picture on	TLBR	L
89.	Not like the	TLBR	L	113.	Longer 3/3 5/6	+ -	
156.	Show picture on	TLBR	R	25.	Find clock	TLBR	R
150.	Show picture on	TLBR	L	6.	Find horse	TLBR	L
69.	People eat most	TLBR	T	141.	35 + 10(2)5	TLBR	L
143.	Show picture on	TLBR	T	58.	Looks just like	TLBR	R
42.	Looks just like	TLBR	R	136.	Whole 1/4	TLBR	L
47.	Looks just like	TLBR	R	153.	Show picture on	TLBR	L
75.	Used W/broom	TLBR	L	147.	Show picture on	TLBR	T
3.	Find axe	TLBR	T	86.	Not eyelid	TLBR	R
15.	Find finished	TLBR	R	11.	Find tree	TLBR	B
62.	From which we pour	TLBR	L	59.	Looks just like	TLBR	T
110.	Not like the	TLBR	T	82.	Biggest if real	TLBR	T
54.	Looks just like	TLBR	L	126.	Find no. which	TLBR	T
55.	Looks just like	TLBR	L	154.	Show picture on	TLBR	B
48.	Looks just like	TLBR	R	64.	Keeps dry	TLBR	T
41.	Looks just like	TLBR	T	8.	Find arm	TLBR	L
20.	Find anchor	TLBR	L	122.	Find no. which	TLBR	L
79.	Carry most people	TLBR	L	114.	Nest - most	TLBR	B
102.	Not like the	TLBR	B	19.	Find padlock	TLBR	R
34.	Looks just like	TLBR	R	152.	Show picture on	TLBR	B
81.	Smallest if real	TLBR	L	76.	Run fastest	TLBR	L
99.	Not like the	TLBR	L	92.	Not like the	TLBR	T
73.	Butterfly	TLBR	B	142.	Show picture on	TLBR	R
129.	3 + 2	TLBR	L	12.	Find shrubbery	TLBR	T
144.	Show picture on	TLBR	T	106.	Not like the	TLBR	B
78.	Heaviest if real	TLBR	B	127.	Find no. which	TLBR	B
61.	Presses clothes	TLBR	T	1.	Find chair	TLBR	R

RAN.

NO _____

65. Used to fight	TLBR	L
33. Looks just like	TLBR	B
35. Looks just like	TLBR	R
36. Looks just like	TLBR	B
108. Not like the	TLBR	R
101. Not like the	TLBR	R
74. Used W/needle	TLBR	L
72. G Y R B	+ -	
96. Not like the	TLBR	L
112. Smallest	+ -	
138. 24 - 18	TLBR	R
124. Find no. which	TLBR	B
49. Looks just like	TLBR	T
100. Not like the	TLBR	T
134. 4/12	TLBR	L
98. Not like the	TLBR	L
28. Find knot	TLBR	T
26. Find coat	TLBR	R
13. Find refreshment	TLBR	B
88. Closest if real	TLBR	L
4. Find farm	TLBR	B
84. Quarter past 8	TLBR	T
46. Looks just like	TLBR	R
119. Eight apples	TLBR	L
22. Find manufacturing	TLBR	T
95. Not like the	TLBR	T
128. Find no. which	TLBR	B
83. Man always	TLBR	B
80. Seal eats	TLBR	L
118. Six apples	TLBR	B
158. Show picture on	TLBR	R
160. Show picture on	TLBR	T
67. Helps us smell	TLBR	T
159. Show picture on	TLBR	R
123. Find no. which	TLBR	L
68. Grows ground	TLBR	B
140. 4 jacks 1/2	TLBR	R
115. Four eggs	TLBR	T
51. Looks just like	TLBR	B
149. Show picture on	TLBR	R
63. Temperature	TLBR	L
56. Looks just like	TLBR	B
38. Looks just like	TLBR	T
32. Find acquiring	TLBR	T
44. Looks just like	TLBR	T
120. 5 3 4 2	+ -	
85. Paper made	TLBR	R
157. Show picture on	TLBR	B
60. Cuts	TLBR	R
93. Not like the	TLBR	L
117. Most apples	TLBR	L
121. Find no. which	TLBR	T
132. 12 - 5	TLBR	B
146. Show picture on	TLBR	L
27. Find retire	TLBR	T
87. House must	TLBR	R
14. Find weapon	TLBR	L
137. Whole 2/3	TLBR	B
30. Find bolt	TLBR	R
135. Whole 1/2	TLBR	R
90. Not like the	TLBR	B
52. Looks just like	TLBR	L
50. Looks just like	TLBR	L
39. Looks just like	TLBR	R
17. Find collision	TLBR	T
18. Find autumn	TLBR	B
31. Find astonishment	TLBR	R
77. Milk people	TLBR	R
116. Three eggs	TLBR	R
37. Looks just like	TLBR	T
107. Not like the	TLBR	B
9. Find sill	TLBR	R
131. 5 - 1 - 1	TLBR	T
57. Looks just like	TLBR	L
111. Biggest	+ -	
109. Not like the	TLBR	B
29. Find plunge	TLBR	L
43. Looks just like	TLBR	L
155. Show picture on	TLBR	L
21. Find nose	TLBR	T

ABSOLUTE ORDER

NO ______

IQ ______

Name ____________________ Date ______________ RS ______

School ____________________ B D ______________ CA ______

Picture Vocabulary

1.	Chair	TLBR R
11.	Tree	TLBR B
2.	Stool	TLBR R
21.	Nose	TLBR T
4.	Farm	TLBR B
15.	Finished	TLBR R
13.	Refreshment	TLBR B
3.	Axe	TLBR T
8.	Arm	TLBR L
7.	Dessert	TLBR R
16.	Uniform	TLBR L
26.	Coat	TLBR R
5.	Pitcher	TLBR L
10.	Cord	TLBR L
6.	Horse	TLBR L
19.	Padlock	TLBR R
14.	Weapon	TLBR L
29.	Plunge	TLBR L
25.	Clock	TLBR R
9.	Sill	TLBR R
12.	Shrubbery	TLBR T
17.	Collision	TLBR T
24.	Ascend	TLBR L
18.	Autumn	TLBR B
22.	Manufacturing	TLBR T
27.	Retire	TLBR T
23.	Swift	TLBR L
30.	Bolt	TLBR R
28.	Knot	TLBR T
31.	Astonishment	TLBR R
32.	Acquiring	TLBR T
20.	Anchor	TLBR L

Info-Compre

61.	Presses clothes	TLBR T
64.	Keeps dry	TLBR T
60.	Cuts	TLBR R
66.	See	TLBR B
68.	Grow ground	TLBR B
78.	Heaviest	TLBR B
65.	Used to fight	TLBR L
72.	G Y R B	+ -
71.	I carry	TLBR B
62.	Pour	TLBR L
67.	Smell	TLBR T
73.	Butterfly	TLBR B
77.	Milk people	TLBR R
63.	Temperature	TLBR L
70.	Man rides	TLBR L
75.	W/broom	TLBR L
74.	W/needle	TLBR L
79.	Carry most	TLBR L
69.	Eat most	TLBR T
85.	Paper made	TLBR R
83.	Man always	TLBR B
80.	Seal eats	TLBR L
81.	Real smallest	TLBR L
86.	Not eyelid	TLBR R
87.	House must	TLBR R
76.	Run fastest	TLBR L
84.	Quarter past 8	TLBR T
82.	Real biggest	TLBR T
88.	Closest	TLBR L

Form Discrimination

33. TLBR B
36. TLBR B
35. TLBR R
34. TLBR R
37. TLBR T
38. TLBR T
39. TLBR R
40. TLBR B
43. TLBR L
41. TLBR T
44. TLBR T
55. TLBR L
50. TLBR L
51. TLBR B
45. TLBR L
54. TLBR L
52. TLBR L
53. TLBR T
46. TLBR R
42. TLBR R
48. TLBR R
56. TLBR B
57. TLBR L
49. TLBR T
47. TLBR R
59. TLBR T
58. TLBR R

ABSOLUTE

NO. _____

Similarities

90. TLBR B
89. TLBR L
91. TLBR T
92. TLBR T
94. TLBR T
95. TLBR T
102. TLBR B
100. TLBR T
93. TLBR L
101. TLBR R
96. TLBR L
99. TLBR L
98. TLBR L
110. TLBR T
97. TLBR R
103. TLBR T
107. TLBR B
104. TLBR T
108. TLBR R
105. TLBR R
109. TLBR B
106. TLBR B

Size and Number

111. Biggest	+ -
113. Longer 3/3 5/6	+ -
112. Smaller	+ -
114. Most	TLBR B
123. Three	TLBR L
122. Three	TLBR L
121. Two	TLBR T
124. Four	TLBR B
116. Three eggs	TLBR R
115. Four eggs	TLBR T
117. Most apples	TLBR L
125. 9 7 8 6	+ -
120. 5 3 4 2	+ -
135. Whole 1/2	TLBR R
118. Six apples	TLBR B
126. Six	TLBR T
119. Eight apples	TLBR L
138. 24 - 18	TLBR R
131. 5 - 1 - 1	TLBR T
129. 3 + 2	TLBR L
130. 4 + 3	TLBR T
128. Eight	TLBR B
137. Whole 2/3	TLBR B
134. 4/12	TLBR L
133. 13 - 8	TLBR T
136. Whole 1/4	TLBR L
132. 12 - 5	TLBR B
139. 12 + 6	TLBR T
127. Eight	TLBR B
141. 35 + 10(2) 5	TLBR L
140. 4 jacks 1/2	TLBR R

Immediate Recall

144. TLBR T
143. TLBR T
142. TLBR R
145. TLBR R
146. TLBR L
147. TLBR T
148. TLBR R
152. TLBR B
149. TLBR R
150. TLBR L
153. TLBR L
155. TLBR L
159. TLBR R
151. TLBR L
154. TLBR B
157. TLBR B
158. TLBR R
156. TLBR R
160. TLBR T

PATTERNED ORDER

NO _____

IQ _____

Name ______________________ Date ______________________ RS _____

School ______________________ B D ______________________ CA _____

111.	Biggest	+ -	
33.	Looks just like	TLBR	B
20.	Find anchor	TLBR	L
36.	Looks just like	TLBR	B
144.	Show picture on	TLBR	T
88.	Closest if real	TLBR	L
1.	Find chair	TLBR	R
143.	Show picture on	TLBR	T
82.	Biggest if real	TLBR	T
90.	Not like the	TLBR	B
142.	Show picture on	TLBR	R
160.	Show picture on	TLBR	T
61.	Presses clothes	TLBR	T
35.	Looks just like	TLBR	R
140.	4 jacks 1/2	TLBR	R
145.	Show picture on	TLBR	R
64.	Keeps dry	TLBR	T
32.	Find acquiring	TLBR	T
113.	Longer 3/3 5/6	+ -	
11.	Find tree	TLBR	B
141.	35 + 10(2)5	TLBR	L
146.	Show picture on	TLBR	L
147.	Show picture on	TLBR	T
127.	Find number which	TLBR	B
60.	Cuts	TLBR	R
34.	Looks just like	TLBR	R
106.	Not like the	TLBR	B
2.	Find stool	TLBR	R
112.	Smallest	+ -	
84.	Quarter past 8	TLBR	T
89.	Not like the	TLBR	L
37.	Looks just like	TLBR	T
139.	12 + 6	TLBR	T
114.	Most eggs	TLBR	B
66.	Helps us see	TLBR	B
31.	Find astonishment	TLBR	R
68.	Grow ground	TLBR	B
21.	Find nose	TLBR	T
132.	12 - 5	TLBR	B
38.	Looks just like	TLBR	T
91.	Not like the	TLBR	T
109.	Not like the	TLBR	B
4.	Find farm	TLBR	B
78.	Heaviest	TLBR	B
105.	Not like the	TLBR	R
148.	Show picture on	TLBR	R
15.	Find finished	TLBR	R
156.	Show picture on	TLBR	R
65.	Used to fight	TLBR	L
72.	G Y R B	+ -	
108.	Not like the	TLBR	R
92.	Not like the	TLBR	T
39.	Looks just like	TLBR	R
136.	Whole 1/4	TLBR	L
94.	Not like the	TLBR	T
123.	Find no. which	TLBR	L
158.	Show picture on	TLBR	R
152.	Show picture on	TLBR	B
122.	Find no. which	TLBR	L
58.	Looks just like	TLBR	R
13.	Find refreshment	TLBR	B
149.	Show picture on	TLBR	R
133.	13 - 8	TLBR	T
121.	Find no. which	TLBR	T
71.	I could carry	TLBR	B
134.	4/12	TLBR	L
124.	Find no. which	TLBR	B
116.	Three eggs	TLBR	R
59.	Looks just like	TLBR	R
3.	Find axe	TLBR	T
62.	Pour	TLBR	L
47.	Looks just like	TLBR	R
67.	Helps us smell	TLBR	T
73.	Butterfly	TLBR	B
49.	Books just like	TLBR	T
95.	Not like the	TLBR	T
77.	Milk people	TLBR	R
137.	Whole 2/3	TLBR	B
40.	Looks just like	TLBR	B
115.	Four eggs	TLBR	T

PAT.

NO. _____

28.	Find knot	TLBR T	126.	Find no. which	TLBR T
63.	Temperature	TLBR L	6.	Find horse	TLBR L
117.	Most apples	TLBR L	107.	Not like the	TLBR B
30.	Find bolt	TLBR R	99.	Not like the	TLBR L
125.	9 7 8 6	+ -	51.	Looks just like	TLBR B
8.	Find arm	TLBR L	128.	Find no. which	TLBR B
76.	Run fastest	TLBR L	19.	Find padlock	TLBR R
7.	Find dessert	TLBR R	98.	Not like the	TLBR L
43.	Looks just like	TLBR L	12.	Find shrubbery	TLBR T
23.	Find swift	TLBR L	119.	Eight apples	TLBR L
120.	5 3 4 2	+ -	14.	Find weapon	TLBR L
135.	Whole 1/2	TLBR R	17.	Find collision	TLBR T
104.	Not like the	TLBR T	138.	24 - 18	TLBR R
102.	Not like the	TLBR B	29.	Find plunge	TLBR L
118.	Six apples	TLBR B	24.	Find ascend	TLBR L
57.	Looks just like	TLBR L	110.	Not like the	TLBR T
150.	Show picture on	TLBR L	153.	Show picture on	TLBR L
16.	Find uniform	TLBR L	80.	Seal eats	TLBR L
27.	Find retire	TLBR T	131.	5 - 1 - 1	TLBR T
100.	Not like the	TLBR T	45.	Looks just like	TLBR L
41.	Looks just like	TLBR T	81.	Smallest if real	TLBR L
56.	Looks just like	TLBR B	25.	Find clock	TLBR R
70.	Man rides	TLBR T	155.	Show picture on	TLBR L
44.	Looks just like	TLBR T	86.	Not eyelid	TLBR R
48.	Looks just like	TLBR R	54.	Looks just like	TLBR L
55.	Looks just like	TLBR L	159.	Show picture on	TLBR R
75.	Used W/broom	TLBR L	157.	Show picture on	TLBR B
22.	Manufacturing	TLBR T	97.	Not like the	TLBR R
50.	Looks just like	TLBR L	151.	Show picture on	TLBR L
26.	Find coat	TLBR R	83.	Man always	TLBR B
42.	Looks just like	TLBR T	79.	Carry most	TLBR L
93.	Not like the	TLBR L	52.	Looks just like	TLBR L
74.	Used W/needle	TLBR L	46.	Looks just like	TLBR R
103.	Not like the	TLBR T	53.	Looks just like	TLBR T
101.	Not like the	TLBR R	129.	3 + 2	TLBR L
5.	Find pitcher	TLBR L	130.	4 + 3	TLBR T
87.	House must	TLBR R	9.	Find sill	TLBR R
10.	Find cord	TLBR L	154.	Show picture on	TLBR B
96.	Not like the	TLBR L	69.	Eat most	TLBR T
18.	Find autumn	TLBR B	85.	Paper made	TLBR R

APPENDIX B

Item Analysis Data for PTI Answer Sheets

Item Analysis Data

Published Order

Item	% Correct	Item	% Correct	Item	% Correct
* 1	93	22	43	43	72
* 2	90	23	40	44	68
3	77	24	46	45	56
4	85	* 25	56	46	48
5	65	* 26	67	47	37
6	63	27	41	48	42
7	73	28	38	* 49	38
8	74	29	60	* 50	67
* 9	57	30	39	51	62
*10	64	31	29	52	52
11	91	32	23	53	52
12	46	* 33	96	54	54
13	80	* 34	90	55	68
14	61	35	92	56	42
15	84	36	95	* 57	41
16	70	37	88	* 58	33
*17	46	38	86	59	36
*18	45	39	82	60	90
19	62	40	76	61	92
20	11	* 41	68	62	77
21	87	* 42	44	63	75

PUB.

Item	% Correct	Item	% Correct	Item	% Correct
64	92	86	47	108	32
*65	84	87	45	109	30
*66	88	88	15	110	57
67	77	* 89	89	111	97
68	88	* 90	93	112	90
69	50	91	86	*113	91
70	68	92	82	*114	88
71	79	93	66	115	76
72	83	94	82	116	77
*73	77	95	76	117	74
*74	66	96	63	118	71
75	67	* 97	52	119	61
76	40	* 98	62	120	72
77	76	99	63	*121	79
78	85	100	69	*122	80
79	52	101	66	123	81
80	47	102	71	124	78
*81	47	103	45	125	74
*82	17	104	41	126	63
83	48	*105	30	127	25
84	28	*106	26	128	46
85	50	107	45	*129	52

PUB.

Item	% Correct	Item	% Correct	Item	% Correct
*130	48	152	81		
131	57	*153	57		
132	30	*154	50		
133	33	155	55		
134	34	156	31		
135	72	157	48		
136	33	158	33		
*137	38	*159	53		
*138	60	*160	19		
139	28				
140	19				
141	23				
142	92				
143	93				
144	95				
*145	92				
*146	91				
147	91				
148	85				
149	80				
150	70				
151	52				

Item Analysis Data

Random Order

Item	% Correct	Item	% Correct	Item	% Correct
*104	41	3	77	40	76
* 94	82	15	84	16	70
103	45	62	77	148	85
71	79	*110	57	105	30
23	40	* 54	54	139	28
10	64	55	68	66	88
70	68	48	42	* 2	90
24	46	41	68	* 45	56
* 7	73	20	11	130	48
* 53	52	79	52	125	74
5	65	102	71	151	52
97	52	* 34	90	113	91
145	92	* 81	47	25	56
89	89	99	63	6	63
156	31	73	67	*141	23
150	70	129	52	* 58	33
* 69	50	144	95	136	33
*143	93	78	85	153	57
42	44	61	92	147	91
47	37	* 91	86	86	47
75	67	*133	33	11	91

RAN.

Item	% Correct	Item	% Correct	Item	% Correct
59	36	101	66	83	48
* 82	17	74	66	80	47
*126	63	72	83	118	71
154	50	* 96	63	158	33
64	92	*112	90	160	19
8	74	138	60	* 67	77
122	80	124	78	*159	53
114	88	49	38	123	81
19	62	100	69	68	88
*152	81	134	34	140	19
* 76	47	98	62	115	76
92	82	* 28	38	51	62
142	92	* 26	67	149	80
12	46	13	80	* 63	75
106	26	88	15	* 56	42
127	25	4	85	38	86
1	93	84	28	32	23
* 65	84	46	48	44	68
* 33	96	119	61	120	72
35	92	* 22	43	85	50
36	95	* 95	76	157	48
108	32	128	46	* 60	90

RAN.

Item	% Correct	Item	% Correct	Item	% Correct
* 93	66	9	51		
117	74	*131	57		
121	79	* 57	41		
132	30	111	97		
146	91	109	30		
27	41	29	60		
87	45	43	72		
* 14	61	*155	55		
*137	38	* 21	87		
30	39				
135	72				
90	93				
52	52				
50	67				
39	82				
* 17	46				
* 18	45				
31	29				
77	76				
116	77				
37	88				
107	45				

Item Analysis Data

Absolute Order

Item	% Correct	Item	% Correct	Item	% Correct
* 1	93	17	46	44	68
* 11	91	24	46	55	68
2	90	18	45	50	67
21	87	* 22	43	51	62
4	85	* 27	41	45	56
15	84	23	40	54	54
13	80	30	39	* 52	52
3	77	28	38	* 53	52
* 8	74	31	29	46	48
* 7	73	32	23	42	44
16	70	20	11	48	42
26	67	* 33	96	56	42
5	65	* 36	95	57	41
10	64	35	92	49	38
6	63	34	90	* 47	37
19	62	37	88	* 59	36
* 14	61	38	86	58	33
* 29	60	39	82	61	92
25	56	40	76	64	92
9	51	* 43	72	60	90
12	46	* 41	68	66	88

ABS.

Item	% Correct	Item	% Correct	Item	% Correct
68	88	84	28	105	30
* 78	85	82	17	109	30
* 65	84	88	15	106	26
72	83	* 90	93	111	97
71	79	* 89	89	113	91
62	77	91	86	*112	90
67	77	92	82	*114	88
73	77	94	82	123	81
77	76	95	76	122	80
* 63	75	102	71	121	79
* 70	68	100	69	124	78
75	67	* 93	66	116	77
74	66	*101	66	115	76
79	52	96	63	*117	74
69	50	99	63	*125	74
85	50	98	62	120	72
83	48	110	57	135	72
* 80	47	97	52	118	71
* 81	47	103	45	126	63
86	47	*107	45	119	61
87	45	*104	41	138	60
76	40	108	32	*131	57

ABS.

Item	% Correct	Item	% Correct	Item	% Correct
*129	52	153	57		
130	48	*155	55		
128	46	*159	53		
137	38	151	52		
134	34	154	50		
133	33	157	48		
136	33	158	33		
*132	30	*156	31		
*139	28	*160	19		
127	25				
141	23				
140	19				
144	95				
143	93				
142	92				
*145	92				
*146	91				
147	91				
148	85				
152	81				
149	80				
150	70				

Item Analysis Data

Patterned Order

Item	% Correct	Item	% Correct	Item	% Correct
*111	97	146	91	4	85
* 33	96	147	91	78	85
20	11	127	25	105	30
36	95	* 60	90	148	85
144	95	* 34	90	15	84
88	15	106	26	156	31
1	93	2	90	* 65	84
143	93	112	90	* 72	83
* 82	17	84	28	108	32
* 90	93	89	89	92	82
142	92	37	88	39	82
160	19	*139	28	136	33
61	92	*114	88	94	82
35	92	66	88	123	81
140	19	31	29	*158	33
145	92	68	88	*152	81
* 64	92	21	87	122	80
* 32	23	132	30	58	33
113	91	38	86	13	80
11	91	* 91	86	149	80
141	23	*109	30	133	33

PAT.

Item	% Correct	Item	% Correct	Item	% Correct
121	79	8	74	22	43
* 71	79	76	40	50	67
*134	34	7	73	26	67
124	78	* 43	72	42	44
116	77	* 23	40	93	66
59	36	120	72	* 74	66
3	77	135	72	*103	45
62	77	104	41	101	66
47	37	102	71	5	65
* 67	77	118	71	87	45
* 73	77	57	41	10	64
49	38	*150	70	96	63
95	76	* 16	70	18	45
77	76	27	41	*126	63
137	38	100	69	* 6	63
40	76	41	68	107	45
115	76	56	42	99	63
* 28	38	70	68	51	62
* 63	75	44	68	128	46
117	74	* 48	42	19	62
30	39	* 55	68	98	62
125	74	75	67	* 12	46

PAT.

Item	% Correct	Item	% Correct	Item	% Correct
*119	61	52	52		
14	61	* 46	48		
17	46	* 53	52		
138	60	129	52		
29	60	130	48		
24	46	9	57		
110	57	154	50		
*153	57	* 69	50		
80	47	* 85	50		
131	57				
45	56				
81	47				
25	56				
155	55				
86	47				
* 54	54				
*159	53				
157	48				
97	52				
151	52				
83	48				
79	52				

Appendix C

PTI Behavioral Rating Scale

PTI

Behavioral Rating Scale

Name ______________________ Date _________ Examiner ________

Observations of behavior

Views all pictures before choosing a response	_ _ _ _ _ _	Frequently selects one (TLBR) position
*Must be told to look at cards	_ _ _ _ _ _	Studies cards carefully
Attention fixed on each card	_ _ _ _ _ _	Attention must be directed to many cards
*Readily says "I don't know"	_ _ _ _ _ _	Attempts to answer each request
Very hesitant in answering	_ _ _ _ _ _	Chooses answers quickly
Sits quietly	_ _ _ _ _ _	Moves about constantly
*Does not talk	_ _ _ _ _ _	Talks freely
Speech intelligible	_ _ _ _ _ _	Speech unintelligible
Superior diction	_ _ _ _ _ _	Poor diction
*Uses single words	_ _ _ _ _ _	Uses sentences of average length

Other: ____

____

____

____

____

____

____

Interpretations of Reactions to the task and to the examiner

Serious	_ _ _ _ _	Playful
*Indifferent	_ _ _ _ _	Enthusiastic
Works with eagerness	_ _ _ _ _	Performs reluctantly or unwillingly
Socially confident	_ _ _ _ _	Shy
*Apprehensive, Fearful	_ _ _ _ _	Does not fear a mistake
Trusting	_ _ _ _ _	Suspicious
Recognizes errors	_ _ _ _ _	Does not recognize failures or errors
*Unaware of inadequate answer	_ _ _ _ _	Overemphasis of inadequacy
Persists when failure is apparent	_ _ _ _ _	Attempts to change from difficult to easy task
Redoubles effort when puzzled	_ _ _ _ _	Gives up easily
*Instructions repeated frequently	_ _ _ _ _	Understands instructions readily
Maintains high interest	_ _ _ _ _	Tires quickly and attention difficult to regain
Considers response carefully	_ _ _ _ _	Reacts impulsively
Consistent effort	_ _ _ _ _	Inconsistent effort

* denotes items scored right to left. Others scored left to right.